GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 6, 2002, 09:38:31 ; Search time 12.18 Seconds
(without alignments)
29.561 Million cell updates/sec

Title: US-09-897-042-18
Perfect score: 85
Sequence: 1 MNLVPMNPLVMNEFMN 16

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 96934

Minimum DB seq length: 0
Maximum DB seq length: 16

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	27	31.8	10	2	US-07-963-538B-14	Sequence 14, Appl
2	27	31.8	11	3	US-08-951-944-8	Sequence 8, Appli
3	27	31.8	15	4	US-08-408-930A-4	Sequence 4, Appli
4	26	30.6	9	3	US-09-075-257A-1	Sequence 1, Appli
5	26	30.6	9	4	US-09-534-639-1	Sequence 1, Appli
6	26	30.6	10	3	US-09-075-257A-10	Sequence 10, Appl
7	26	30.6	10	4	US-09-534-639-10	Sequence 10, Appl
8	26	30.6	13	1	US-08-128-971B-2	Sequence 2, Appli
9	26	30.6	15	4	US-08-256-104-5	Sequence 5, Appli
10	26	30.6	16	4	US-08-602-999A-172	Sequence 172, App
11	25	29.4	14	1	US-08-160-317-4	Sequence 4, Appli
12	25	29.4	14	1	US-08-351-147-4	Sequence 4, Appli
13	25	29.4	14	1	US-08-471-154-4	Sequence 4, Appli
14	25	29.4	16	1	US-08-574-763-7	Sequence 7, Appli
15	25	29.4	16	4	US-08-602-999A-234	Sequence 234, App
16	25	29.4	16	5	PCT-US91-09422-31	Sequence 31, Appl
17	24.5	28.8	12	3	US-08-951-944-10	Sequence 10, Appl
18	24	28.2	9	4	US-09-177-249-299	Sequence 299, App
19	24	28.2	10	3	US-08-159-339A-474	Sequence 474, App
20	24	28.2	10	3	US-08-159-339A-475	Sequence 475, App
21	24	28.2	10	4	US-08-685-558A-2	Sequence 2, Appli
22	24	28.2	13	1	US-08-464-531-106	Sequence 106, App
23	24	28.2	13	2	US-08-461-598-106	Sequence 106, App
24	24	28.2	13	3	US-08-322-137-106	Sequence 106, App
25	24	28.2	13	4	US-08-582-333A-42	Sequence 42, Appl
26	24	28.2	14	1	US-07-991-199D-5	Sequence 5, Appli
27	24	28.2	14	3	US-09-082-039A-12	Sequence 12, Appl
28	24	28.2	14	5	PCT-US93-12246-5	Sequence 5, Appli
29	24	28.2	15	1	US-08-221-583-61	Sequence 61, Appl
30	24	28.2	15	1	US-08-221-583-62	Sequence 62, Appl
31	24	28.2	15	1	US-07-965-667A-19	Sequence 19, Appl
32	24	28.2	15	3	US-08-484-819-19	Sequence 19, Appl
33	24	28.2	15	4	US-08-602-999A-171	Sequence 171, App
34	24	28.2	15	5	PCT-US93-10197-19	Sequence 19, Appl
35	24	28.2	15	5	PCT-US95-04018-61	Sequence 61, Appl
36	24	28.2	15	5	PCT-US95-04018-62	Sequence 62, Appl
37	23	27.1	7	2	US-08-632-598-11	Sequence 11, Appl
38	23	27.1	7	4	US-09-231-240-11	Sequence 11, Appl
39	23	27.1	8	1	US-08-636-386-4	Sequence 4, Appli
40	23	27.1	8	5	PCT-US95-13848-4	Sequence 4, Appli
41	23	27.1	9	2	US-08-318-856A-53	Sequence 53, Appl
42	23	27.1	9	4	US-09-518-046-110	Sequence 110, App
43	23	27.1	10	1	US-08-636-386-5	Sequence 5, Appli
44	23	27.1	10	1	US-08-408-604A-71	Sequence 71, Appl
45	23	27.1	10	5	PCT-US95-13848-5	Sequence 5, Appli
46	23	27.1	13	1	US-08-189-772-2	Sequence 2, Appli
47	23	27.1	13	1	US-08-188-277B-11	Sequence 11, Appl
48	23	27.1	13	2	US-08-467-046-7	Sequence 7, Appli
49	23	27.1	14	1	US-08-188-277B-24	Sequence 24, Appl
50	23	27.1	15	1	US-08-449-207-4	Sequence 4, Appli

ALIGNMENTS

```
RESULT   1
US-07-963-538B-14
; Sequence 14, Application US/07963538B
; Patent No. 5851983
;  GENERAL INFORMATION:
;    APPLICANT:  SUGIYAMA, TAKASHI
;    APPLICANT:  KAMIMURA, TAKASHI
;    APPLICANT:  MASUDA, KENICHI
;    APPLICANT:  OKADA, MASAHIRO
;    APPLICANT:  OHTSUKA, EIKO
;    APPLICANT:  IMAIZUMI, ATSUSHI
;    APPLICANT:  WATANABE, KUNIHITO
;    APPLICANT:  SUGA, TETSUYA
;    APPLICANT:  MATSUMOTO, YOHICHI
;    APPLICANT:  TAKEUCHI, AKIKO
;    TITLE OF INVENTION:  ELASTASE INHIBITORY POLYPEPTIDE AND
;    TITLE OF INVENTION:  PROCESS FOR PRODUCTION THEREOF BY RECOMBINANT GENE
;    TITLE OF INVENTION:  TECHNOLOGY
;    NUMBER OF SEQUENCES:  36
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  COOLEY GODWARD LLP
;      STREET:  FIVE PALO ALTO SQUARE, 4TH FLOOR
;      CITY:  PALO ALTO
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94306-2155
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/963,538B
;      FILING DATE:  20-OCT-1992
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/843,359
;      FILING DATE:  25-FEB-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/408,483
;      FILING DATE:  22-AUG-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 4-212399
;      FILING DATE:  17-JUL-1992
;    PRIOR APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  JP 4-212398
;      FILING DATE:  17-JUL-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 3-355553
;      FILING DATE:  24-DEC-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 62-330219
;      FILING DATE:  28-DEC-1987
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  NEELEY PH.D., RICHARD L.
;      REGISTRATION NUMBER:  30,092
;      REFERENCE/DOCKET NUMBER:  TEJN-005/02US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-843-5070
;      TELEFAX:  415-857-0663
;      TELEX:  380816 COOLEY PA
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-07-963-538B-14


  Query Match             31.8%;  Score 27;  DB 2;  Length 10;
  Best Local Similarity   83.3%;  Pred. No. 67;
  Matches    5;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 LVPMNP 8
           ||| ||
Db       4 LVPRNP 9


RESULT   2
US-08-951-944-8
; Sequence 8, Application US/08951944
; Patent No. 6037449
;  GENERAL INFORMATION:
;    APPLICANT:  Zhou, Hai-Yan
;    APPLICANT:  Salih, Erdjan
;    APPLICANT:  Glimcher, Melvin J.
;    TITLE OF INVENTION:  OSTEOMETRIN RELATED PEPTIDES AND
;    TITLE OF INVENTION:  NUCLEIC ACIDS AND DIAGNOSIS AND THERAPIES USING THEM
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows 95
;      SOFTWARE:  FastSEQ for Windows Version 2.0b
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/951,944
;      FILING DATE:  17-OCT-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/028,682
;      FILING DATE:  18-OCT-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Freeman, John W.
;      REGISTRATION NUMBER:  29,066
;      REFERENCE/DOCKET NUMBER:  00108/125001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-5070
;      TELEFAX:  617/542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-951-944-8


  Query Match             31.8%;  Score 27;  DB 3;  Length 11;
  Best Local Similarity   62.5%;  Pred. No. 75;
  Matches    5;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       5 PMNPLVMN 12
           | ||  ||
Db       3 PYNPQYMN 10


RESULT   3
US-08-408-930A-4
; Sequence 4, Application US/08408930A
; Patent No. 6242567
;  GENERAL INFORMATION:
;    APPLICANT:  Hema Pande, Arthur D. Riggs, John
;    APPLICANT:  A. Zaia and Brian R. Clark
;    TITLE OF INVENTION:  Method For Detection and
;    TITLE OF INVENTION:  Prevention of Human Cytomegalovirus Infection
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  City of Hope
;      STREET:  1500 East Duarte Road
;      CITY:  Duarte
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  91010-0269
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3M Double Density
;      MEDIUM TYPE:  5 1/4" diskette
;      COMPUTER:  Wang PC
;      OPERATING SYSTEM:  MS DOS Version 3.20
;      SOFTWARE:  WordPerfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/408,930A
;      FILING DATE:  22 March 1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/978,151
;      FILING DATE:  17 No. 6242567ember 1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Irons, Edward S.
;      REGISTRATION NUMBER:  16,541
;      REFERENCE/DOCKET NUMBER:  Pande
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 626-3564
;      TELEFAX:  (202) 783-6031
;      TELEX:  No. 6242567e
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  15
;      TYPE:  Amino Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Unknown
US-08-408-930A-4


  Query Match             31.8%;  Score 27;  DB 4;  Length 15;
  Best Local Similarity   50.0%;  Pred. No. 1.1e+02;
  Matches    6;  Conservative    1;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       2 NLVPMNPLVMNE 13
           |||||   |  :
Db       1 NLVPMVATVQGQ 12


RESULT   4
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 6, 2002, 09:36:16 ; Search time 25.16 Seconds
(without alignments)
47.105 Million cell updates/sec

Title: US-09-897-042-18
Perfect score: 85
Sequence: 1 MNLVPMNPLVMNEFMN 16

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 178144

Minimum DB seq length: 0
Maximum DB seq length: 16

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : A_Geneseq_1101:*
1: /SIDS2/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	31	36.5	15	17	AAR97927	Japan cedar pollen
2	28	32.9	11	19	AAW83283	NPF motif EH domai
3	27	31.8	9	17	AAW00715	Human Stat 5 fragm
4	27	31.8	9	22	AAM07384	HLA-A1 nonamer #91
5	27	31.8	9	22	AAM07797	HLA-A *0201 noname
6	27	31.8	9	22	AAM08475	HLA-B *2705 noname
7	27	31.8	9	22	AAM09077	HLA-B *0702 noname
8	27	31.8	9	22	AAM11238	HLA-A26 nonamer #1
9	27	31.8	9	22	AAM12037	HLA-B8 nonamer #27
10	27	31.8	9	22	AAM12642	HLA-B *1510 noname
11	27	31.8	10	22	AAM09183	HLA-A *0201 decame
12	27	31.8	10	22	AAM09603	HLA-A1 decamer #17
13	27	31.8	10	22	AAM09633	HLA-A1 decamer #20
14	27	31.8	10	22	AAM09647	HLA-B *0702 decame
15	27	31.8	10	22	AAM10021	HLA-B *0702 decame
16	27	31.8	10	22	AAM10974	HLA-A *0201 decame
17	27	31.8	10	22	AAM12823	HLA-A26 decamer #6
18	27	31.8	10	22	AAM13449	HLA-A26 decamer #5
19	27	31.8	14	22	AAG98147	Human SNP associat
20	27	31.8	15	19	AAW62870	Epitope of the pp6
21	27	31.8	15	20	AAY31203	Human cytomegalovi
22	27	31.8	15	22	AAB84955	HCMV 64 kilodalton
23	26	30.6	8	22	AAM09332	HLA-B8 octamer #35
24	26	30.6	8	22	AAM09520	HLA-B8 octamer #79
25	26	30.6	8	22	AAM10247	HLA-B8 octamer #26
26	26	30.6	8	22	AAM10943	HLA-B8 octamer #42
27	26	30.6	9	19	AAW62869	Epitope of the pp6
28	26	30.6	9	19	AAW48979	Nonapeptide immuno
29	26	30.6	9	20	AAY31196	Human cytomegalovi
30	26	30.6	9	20	AAY31197	Human cytomegalovi
31	26	30.6	9	20	AAY31198	Human cytomegalovi
32	26	30.6	9	20	AAY09319	Immunogenic peptid
33	26	30.6	9	21	AAB12406	Immunogenic epitop
34	26	30.6	9	21	AAY54994	HLA-A2 restricted
35	26	30.6	9	22	AAM07208	HLA-B *0702 noname
36	26	30.6	9	22	AAM07232	HLA-B *0702 noname
37	26	30.6	9	22	AAM07249	HLA-B *0702 noname
38	26	30.6	9	22	AAM07321	HLA-A *0201 noname
39	26	30.6	9	22	AAM07427	HLA-A1 nonamer #13
40	26	30.6	9	22	AAM07632	HLA-A1 nonamer #16
41	26	30.6	9	22	AAM07749	HLA-B *2705 noname
42	26	30.6	9	22	AAM07906	HLA-B *0702 noname
43	26	30.6	9	22	AAM08008	HLA-B *2705 noname
44	26	30.6	9	22	AAM08057	HLA-A *0201 noname
45	26	30.6	9	22	AAM08264	HLA-A *0201 noname
46	26	30.6	9	22	AAM08476	HLA-B *2705 noname
47	26	30.6	9	22	AAM08532	HLA-A *0201 noname
48	26	30.6	9	22	AAM08714	HLA-B *2705 noname
49	26	30.6	9	22	AAM08820	HLA-A1 nonamer #46
50	26	30.6	9	22	AAM08821	HLA-A1 nonamer #46

ALIGNMENTS

RESULT 1
AAR97927
ID AAR97927 standard; peptide; 15 AA.
XX
AC AAR97927;
XX
DT 16-AUG-1996 (first entry)
XX
DE Japan cedar pollen mature allergen Cry j II amino acids 281-295.
XX
KW Allergen; epitope; overlapping peptide; Cry j II; cedar pollen;
KW Sugi pollinosis; diagnosis; treatment.
XX
OS Cryptomeria japonica.
XX
PN JP08047392-A.
XX
PD 20-FEB-1996.
XX
PF 07-NOV-1994; 94JP-0297840.
XX
PR 26-MAY-1994; 94JP-0134868.
PR 05-NOV-1993; 93JP-0276773.
XX
PA (MEIP) MEIJI MILK PROD CO LTD.
XX
DR WPI; 1996-166249/17.
XX
PT Japan cedar pollen allergen Cry j II epitope - comprises at least

PT part of specified 460 amino acid protein
XX
PS Disclosure; Fig 5; 17pp; Japanese.
XX
CC AAR97871-R97960 are overlapping peptides used for the epitope mapping
CC of the Japan cedar pollen allergen Cry j II. Cry j II and allergenic
CC peptides of it are useful in the diagnosis, prevention and treatment
CC of Sugi pollinosis, the allergic reaction to Japan cedar pollen.
CC Significant regions of the allergen were identified using the
CC overlapping peptides of the full epitope derived from a Cry j II
CC antigen-specific T cell line. Amino acids 66-80 (AAR97884) and 186-200
CC (R978908) of the full mature 460 amino acid allergen are the most
CC allergenic of the 90 peptides tested.
XX
SQ Sequence 15 AA;

Query Match 36.5%; Score 31; DB 17; Length 15;
Best Local Similarity 50.0%; Pred. No. 41;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 7 NPLVMNEF 14
||:::|:|
Db 4 npilinqf 11

RESULT 2
AAW83283
ID AAW83283 standard; peptide; 11 AA.
XX
AC AAW83283;
XX
DT 08-FEB-1999 (first entry)
XX
DE NPF motif EH domain binding peptide #64.
XX
KW Human; cytoplasmic protein; EH-containing protein; eps15; eps15R;
KW eps15 homology; intracellular interaction; EH domain binding specificity;
KW signal transducer; NPF motif; h-NUMB; h-NUMB-R; h-RAB; h-RAB-R; ehb3;
KW ehb10; ehb21; cell proliferation; diagnosis; detection.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO9846744-A1.
XX
PD 22-OCT-1998.
XX
PF 06-APR-1998; 98WO-IT00077.
XX
PR 15-APR-1997; 97IT-MI00868.
XX
PA (EUON-) IST EURO DI ONCOLOGIA SRL.
XX
PI Di Fiore PP, Doria M, Pelicci PG, Salcini AE;
XX
DR WPI; 1998-594574/50.
XX
PT New isolated EH domain binding proteins and peptide(s) - obtained
PT using signal transducers eps15 and eps15R containing EH domains by
PT detecting specific binding activity.
XX
PS Claim 11; Page 73; 90pp; English.
XX
CC The present invention describes a new intracellular interactor and novel
CC protein:protein interaction (EH) domain binding protein having (parts
CC of) one of the following human derived sequences designated: (I) h-NUMB;
CC (II) h-NUMB-R; (III) h-RAB-R; (IV) ehb3; (V) ehb10; or (VI) ehb21. The
CC present invention also describes peptides containing at least one NPF
CC (Asp-Pro-Phe) motif, able to bind to a protein with at least one EH
CC domain. The proteins with a NPF-containing peptide or the peptides
CC themselves can be used to identify and purify EH containing proteins.
CC [illegible]ense RNA, complementary to mRNA encoding h-NUMB, h-NUMB-R, h-RAB-R,
CC ehb3, ehb10, or ehb21, can be used for diagnostic and therapeutic uses.
CC Products from the present invention can also be used to develop agents
CC for use in control of cell proliferation. AAW83220 to AAW83284 represent
CC NPF motif EH domain binding peptides.
XX
SQ Sequence 11 AA;

Query Match 32.9%; Score 28; DB 19; Length 11;
Best Local Similarity 44.4%; Pred. No. 97;
Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 5 PMNPLVMNE 13
|:|| :|:
Db 3 pvnpfqvnq 11

RESULT 3
AAW00715
ID AAW00715 standard; Peptide; 9 AA.
XX
AC AAW00715;
XX
DT 06-DEC-1996 (first entry)
XX
DE Human Stat 5 fragment (aa735-743).
XX
KW Signal transducer and activator of transcription 5; Stat 5;
KW interleukin-2; signal transduction; cell proliferation;
KW immune disorder; therapy.
XX
OS Homo sapiens.
XX
PN WO9626292-A1.
XX
PD 29-AUG-1996.
XX
PF 22-FEB-1996; 96WO-US02457.
XX
PR 23-FEB-1995; 95US-0393333.
XX
PA (TULA-) TULARIK INC.
XX
PI Hou J, McKnight SL, Schindler U;
XX
DR WPI; 1996-402382/40.
XX
PT Human signal transducer and activator of transcription 5 protein -
PT used for treating cellular proliferation disorders, pref. of immune
PT cells
XX
PS Claim 2; Page 36; 42pp; English.
XX
CC Fragments (AAW00707-20) of human signal transducer and activator of
CC transcription 5 (hStat 5) (AAW00706) retain hStat 5-specific
CC binding affinity and are capable of eliciting antibodies capable
CC of distinguishing hStat 5 from other Stats and Stat-related
CC proteins such as mammary gland factor. They can be produced
CC by chemical synthesis or by recombinant methods. They are
CC useful e.g. in two-hybrid screening to identify intracellular
CC targets which specifically bind to hStat 5. Screening assays
CC incorporating the fragments are used to identify agents useful
CC in the diagnosis, prognosis or treatment of diseases associated
CC with undesirable cell growth, differentiation and/or cytokine
CC signal responsiveness.
XX
SQ Sequence 9 AA;

Query Match 31.8%; Score 27; DB 17; Length 9;
Best Local Similarity 57.1%; Pred. No. 4.3e+05;
Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 6, 2002, 09:40:11 ; Search time 24.51 Seconds
(without alignments)
95.486 Million cell updates/sec

Title: US-09-897-042-18
Perfect score: 85
Sequence: 1 MNLVPMNPLVMNEFMN 16

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 3030

Minimum DB seq length: 0
Maximum DB seq length: 16

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : SPTREMBL_17:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	35	41.2	16	11	Q9QVM7	Q9qvm7 rattus sp.
2	25	29.4	15	11	O88175	O88175 mus musculu
3	24	28.2	15	13	Q9PTM6	Q9ptm6 poephila gu
4	24	28.2	16	8	Q36789	Q36789 solanum nig
5	23	27.1	15	2	Q54325	Q54325 staphylococ
6	22.5	26.5	16	10	P93232	P93232 lycopersico
7	22	25.9	8	5	P82618	P82618 periplaneta
8	22	25.9	8	8	Q9MD43	Q9md43 rattus norv
9	22	25.9	13	8	Q33417	Q33417 digitalis p
10	22	25.9	15	2	Q9R521	Q9r521 francisella
11	22	25.9	15	8	Q37016	Q37016 nicotiana a
12	22	25.9	16	5	Q9TWN7	Q9twn7 toxoplasma
13	21	24.7	9	10	Q9S8J8	Q9s8j8 oryza sativ
14	21	24.7	13	4	Q9UPB6	Q9upb6 homo sapien
15	21	24.7	13	13	P82866	P82866 rana pipien
16	21	24.7	14	2	Q9R457	Q9r457 clostridium
17	21	24.7	16	1	Q9UWK4	Q9uwk4 methanobact
18	20.5	24.1	9	12	Q67605	Q67605 squash leaf
19	20.5	24.1	9	12	Q67606	Q67606 squash leaf
20	20	23.5	9	12	Q88612	Q88612 middelburg
21	20	23.5	12	6	Q9TQW3	Q9tqw3 bos taurus
22	20	23.5	14	2	Q9KEL4	Q9kel4 bacillus ha
23	20	23.5	14	4	Q9P2A2	Q9p2a2 homo sapien
24	20	23.5	15	4	Q9UCL7	Q9ucl7 homo sapien
25	20	23.5	15	7	Q9TNQ3	Q9tnq3 homo sapien
26	20	23.5	15	11	Q9QV62	Q9qv62 mus sp. dif
27	19.5	22.9	15	2	Q46456	Q46456 clostridium
28	19	22.4	11	4	Q9UCP2	Q9ucp2 homo sapien
29	19	22.4	15	2	O05463	O05463 nitrosospir
30	19	22.4	15	6	Q9TRN8	Q9trn8 sus scrofa
31	19	22.4	15	7	Q9TNP0	Q9tnp0 mus sp. mhc
32	19	22.4	15	13	Q9PRZ0	Q9prz0 bothrops ja
33	19	22.4	16	2	Q9AH61	Q9ah61 neisseria m
34	19	22.4	16	6	Q9TR60	Q9tr60 bos taurus
35	18	21.2	8	4	Q9HCQ0	Q9hcq0 homo sapien
36	18	21.2	10	12	Q84140	Q84140 influenza a
37	18	21.2	11	2	Q9R4B1	Q9r4b1 streptococc
38	18	21.2	12	10	P82441	P82441 nicotiana t
39	18	21.2	12	12	Q84038	Q84038 influenza a
40	18	21.2	13	4	Q16007	Q16007 homo sapien
41	18	21.2	13	10	Q43174	Q43174 solanum tub
42	18	21.2	13	11	Q60517	Q60517 mus musculu
43	18	21.2	13	12	Q83171	Q83171 cauliflower
44	18	21.2	13	12	Q9WMG5	Q9wmg5 sigma virus
45	18	21.2	14	2	O85527	O85527 chlamydia t
46	18	21.2	14	4	O95179	O95179 homo sapien
47	18	21.2	15	2	O68425	O68425 buchnera ap
48	18	21.2	15	4	Q9UMT6	Q9umt6 homo sapien
49	18	21.2	15	4	Q9UC60	Q9uc60 homo sapien
50	18	21.2	15	4	Q9UC45	Q9uc45 homo sapien

ALIGNMENTS

RESULT 1
Q9QVM7
ID Q9QVM7 PRELIMINARY; PRT; 16 AA.
AC Q9QVM7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE ANKYRIN-BINDING GLYCOPROTEIN (FRAGMENT).
OS Rattus sp.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10118;
RN [1]
RP SEQUENCE.
RX MEDLINE=93209979; PubMed=8458865;
RA Davis J.Q., McLaughlin T., Bennett V.;
RT "Ankyrin-binding proteins related to nervous system cell adhesion
RT molecules: candidates to provide transmembrane and intercellular
RT connections in adult brain.";
RL J. Cell Biol. 121:121-133(1993).
SQ SEQUENCE 16 AA; 1825 MW; 201C85AC575E2E9B CRC64;

Query Match 41.2%; Score 35; DB 11; Length 16;
Best Local Similarity 50.0%; Pred. No. 19;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

```
Qy       4 VPMNPLVMNE 13
           :||:| : ||
Db       3 IPMDPSIQNE 12
```

RESULT 2
O88175
ID O88175 PRELIMINARY; PRT; 15 AA.
AC O88175;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)

```
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   NEURAL CELL ADHESION MOLECULE (FRAGMENT).
GN   NCAM.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB-C; TISSUE=LIVER;
RX   MEDLINE=98250618; PubMed=9582442;
RA   Kawahigashi H., Harada Y., Asano A., Nakamura M.;
RT   "A cis-acting regulatory element that affects the alternative splicing
RT   of a muscle-specific exon in the mouse NCAM gene.";
RL   Biochim. Biophys. Acta 1397:305-315(1998).
DR   EMBL; AB001873; BAA31274.1; -.
DR   MGD; MGI:97281; Ncam.
FT   NON_TER       1      1
FT   NON_TER      15     15
SQ   SEQUENCE   15 AA;  1481 MW;  41868EF6117732C2 CRC64;

  Query Match             29.4%;  Score 25;  DB 11;  Length 15;
  Best Local Similarity   66.7%;  Pred. No. 8.8e+02;
  Matches    4;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 LVPMNP 8
           |||::|
Db       8 LVPLSP 13


RESULT   3
Q9PTM6
ID   Q9PTM6      PRELIMINARY;      PRT;    15 AA.
AC   Q9PTM6;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   AROMATASE (FRAGMENT).
GN   AROMATASE.
OS   Poephila guttata (Zebra finch) (Taeniopygia guttata).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Passeriformes; Estrildidae;
OC   Estrildinae; Taeniopygia.
OX   NCBI_TaxID=59729;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20033564; PubMed=10564828;
RA   Ramachandran B., Schlinger B.A., Arnold A.P., Campagnoni A.T.;
RT   "Zebra finch aromatase gene expression is regulated in the brain
RT   through an alternate promoter.";
RL   Gene 240:209-216(1999).
DR   EMBL; AF170273; AAF20041.1; -.
FT   NON_TER      15     15
SQ   SEQUENCE   15 AA;  1745 MW;  29099C05146F360E CRC64;

  Query Match             28.2%;  Score 24;  DB 13;  Length 15;
  Best Local Similarity   50.0%;  Pred. No. 1.3e+03;
  Matches    6;  Conservative    1;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 MNLVPMNPLVMN 12
           | |  :|||  |
Db       1 MVLETLNPLHYN 12


RESULT   4
Q36789
ID   Q36789      PRELIMINARY;   PRT;   16 AA.
AC   Q36789;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   RPS19' PROTEIN (FRAGMENT).
GN   RPS19'.
OS   Solanum nigrum (Black nightshade).
OG   Chloroplast.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX   NCBI_TaxID=4112;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96397499; PubMed=8804393;
RA   Goulding S.E., Olmstead R.G., Morden C.W., Wolfe K.H.;
RT   "Ebb and flow of the chloroplast inverted repeat.";
RL   Mol. Gen. Genet. 252:195-206(1996).
DR   EMBL; Z71249; CAA94964.1; -.
KW   Chloroplast.
FT   NON_TER      16     16
SQ   SEQUENCE   16 AA;  1869 MW;  1AB9DE1D775A138C CRC64;

  Query Match             28.2%;  Score 24;  DB 8;  Length 16;
  Best Local Similarity   44.4%;  Pred. No. 1.4e+03;
  Matches    4;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       7 NPLVMNEFM 15
           || | |  :
Db       8 NPFVANHLL 16


RESULT   5
Q54325
ID   Q54325      PRELIMINARY;      PRT;    15 AA.
AC   Q54325;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   CHLORAMPHENICOL ACETYLTRANSFERASE (FRAGMENT).
GN   CAT.
OS   Staphylococcus intermedius.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Staphylococcus.
OX   NCBI_TaxID=1285;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PCW6;
RX   MEDLINE=96379895; PubMed=8787908;
RA   Lodder G., Schwarz S., Gregory P., Dyke K.;
RT   "Tandem duplication in ermC translational attenuator of the macrolide-
RT   lincosamide-streptogramin B resistance plasmid pSES6 from
RT   Staphylococcus equorum.";
RL   Antimicrob. Agents Chemother. 40:215-217(1996).
DR   EMBL; X82666; CAA57981.1; -.
KW   Transferase.
FT   NON_TER       1      1
SQ   SEQUENCE   15 AA;  1931 MW;  173A6AA93999A8D9 CRC64;

  Query Match             27.1%;  Score 23;  DB 2;  Length 15;
  Best Local Similarity   66.7%;  Pred. No. 1.9e+03;
  Matches    4;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      11 MNEFMN 16
           |||| :
Db       1 MNEFQD 6


RESULT   6
P93232
ID   P93232      PRELIMINARY;      PRT;    16 AA.
AC   P93232;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 6, 2002, 09:40:42 ; Search time 9.79 Seconds
(without alignments)
59.922 Million cell updates/sec

Title: US-09-897-042-18
Perfect score: 85
Sequence: 1 MNLVPMNPLVMNEFMN 16

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 780

Minimum DB seq length: 0
Maximum DB seq length: 16

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID		Description
1	28	32.9	15	1	UBL1_MONDO	P50103	monodelphis
2	22.5	26.5	15	1	PC20_BRANA	P81096	brassica na
3	22	25.9	12	1	TM2A_METMA	P80652	methanosarc
4	20.5	24.1	15	1	UC29_MAIZE	P80635	zea mays (m
5	20	23.5	9	1	UPA7_HUMAN	P30093	homo sapien
6	19	22.4	8	1	FAR8_CALVO	P41863	calliphora
7	19	22.4	11	1	FAR9_CALVO	P41864	calliphora
8	19	22.4	14	1	ATP6_SPIOL	P80086	spinacia ol
9	18.5	21.8	11	1	TKN5_PSEGU	P42990	pseudophryn
10	18.5	21.8	13	1	PHGR_EUBOX	P57793	eubacterium
11	18	21.2	7	1	FAR1_ASCSU	P31889	ascaris suu
12	18	21.2	11	1	TKN3_PSEGU	P42988	pseudophryn
13	18	21.2	12	1	DCML_PSECA	P19919	pseudomonas
14	18	21.2	13	1	CRBL_VESCR	P01518	vespa crabr
15	18	21.2	13	1	LMT4_LOCMI	P41490	locusta mig
16	18	21.2	14	1	CRBL_VESOR	P17236	vespa orien
17	18	21.2	14	1	UC34_MAIZE	P80640	zea mays (m
18	18	21.2	15	1	CYSK_CLOPA	P81340	clostridium
19	18	21.2	15	1	ESTB_SCHGA	P81011	schizaphis
20	18	21.2	15	1	UC13_MAIZE	P80619	zea mays (m
21	18	21.2	16	1	LPK1_LOCMI	P20404	locusta mig
22	17	20.0	6	1	VP19_HSV1K	P23210	herpes simp
23	17	20.0	11	1	CSI5_BACSU	P81095	bacillus su
24	17	20.0	11	1	PQQC_PSEFL	P55173	pseudomonas
25	17	20.0	15	1	ATP2_SPIOL	P80083	spinacia ol
26	17	20.0	15	1	RBS_PHYPA	P80657	physcomitre
27	17	20.0	15	1	RM12_YEAST	P36522	saccharomyc
28	17	20.0	15	1	UE15_HORVU	P34938	hordeum vul
29	17	20.0	16	1	CXA2_CONMA	P56636	conus magus
30	16.5	19.4	13	1	CRBL_ICASP	P17237	icaria sp.
31	16	18.8	9	1	BUK_CLOPA	P81337	clostridium
32	16	18.8	10	1	RRPL_PHODV	P35946	phocine dis
33	16	18.8	10	1	URE3_MORMO	P17339	morganella
34	16	18.8	11	1	RRPL_CHAV	P13179	chandipura
35	16	18.8	13	1	ACT7_SOYBN	P15987	glycine max
36	16	18.8	13	1	CRTC_RANES	P31832	rana escule
37	16	18.8	13	1	TEMD_RANTE	P56919	rana tempor
38	16	18.8	13	1	YPNP_PHOLU	P41122	photorhabdu
39	16	18.8	15	1	COXJ_THUOB	P80979	thunnus obe
40	16	18.8	15	1	UC17_MAIZE	P80623	zea mays (m
41	15	17.6	7	1	MYOM_APLCA	P15513	aplysia cal
42	15	17.6	9	1	MGMT_BOVIN	P29177	bos taurus
43	15	17.6	9	1	UPA6_HUMAN	P30092	homo sapien
44	15	17.6	10	1	COXK_ONCMY	P80332	oncorhynchu
45	15	17.6	10	1	NS1_MYCTU	P81135	mycobacteri
46	15	17.6	10	1	UPA9_HUMAN	P30095	homo sapien
47	15	17.6	11	1	MLG_THETS	P41989	theromyzon
48	15	17.6	11	1	PVK1_PERAM	P41837	periplaneta
49	15	17.6	11	1	UN05_CLOPA	P81350	clostridium
50	15	17.6	12	1	PA2B_VIPBO	P31859	vipera beru

ALIGNMENTS

RESULT 1
UBL1_MONDO
ID UBL1_MONDO STANDARD; PRT; 15 AA.
AC P50103;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE UBIQUITIN CARBOXYL-TERMINAL HYDROLASE ISOZYME L1 (EC 3.4.19.12) (UCH-
DE L1) (UBIQUITIN THIOLESTERASE L1) (PGP 9.5) (FRAGMENT).
GN UCHL1.
OS Monodelphis domestica (Short-tailed grey opossum).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Didelphimorphia; Didelphidae; Monodelphis.
OX NCBI_TaxID=13616;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=96102916; PubMed=8522974;
RA Mann D.A., Trowern A.R., Lavender F.L., Whittaker P.A.,
RA Thompson R.J.;
RT "Identification of evolutionary conserved regulatory sequences in the
RT 5' untranscribed region of the neural-specific ubiquitin C-terminal
RT hydrolase (PGP9.5) gene.";
RL J. Neurochem. 66:35-46(1996).
CC -!- FUNCTION: UBIQUITIN-PROTEIN HYDROLASE IS INVOLVED BOTH IN THE
CC PROCESSING OF UBIQUITIN PRECURSORS AND OF UBIQUINATED PROTEINS.
CC THIS ENZYME IS A THIOL PROTEASE THAT RECOGNIZE AND HYDROLYZE
CC A PEPTIDE BOND AT THE C-TERMINAL GLYCINE OF UBIQUITIN.
CC -!- CATALYTIC ACTIVITY: UBIQUITIN C-TERMINAL THIOLESTER + H(2)O =
CC UBIQUITIN + A THIOL.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C12; ALSO KNOWN AS
CC FAMILY 1 OF UBIQUITIN CARBOXYL-TERMINAL HYDROLASES.
CC ---

CC ---
DR EMBL; U32208; AAA89059.1; -.
DR InterPro; IPR001578; UCH.
DR PROSITE; PS00140; UCH_1; PARTIAL.
KW Ubiquitin conjugation; Hydrolase; Thiol protease; Multigene family.
FT NON_TER 15 15
SQ SEQUENCE 15 AA; 1816 MW; 1B7A3B6E623F6E79 CRC64;

Query Match 32.9%; Score 28; DB 1; Length 15;

```
  Best Local Similarity   46.7%;  Pred. No. 60;
  Matches    7;  Conservative    3;  Mismatches    3;  Indels    2;  Gaps    1;

Qy       1 MNLVPM--NPLVMNE 13
           | | ||  || ::|:
Db       1 MQLKPMEINPEMLNK 15


RESULT   2
PC20_BRANA
ID   PC20_BRANA     STANDARD;      PRT;    15 AA.
AC   P81096;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   20 KDA POLLEN COAT PROTEIN (FRAGMENT).
OS   Brassica napus (Rape).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Brassica.
OX   NCBI_TaxID=3708;
RN   [1]
RP   SEQUENCE.
RC   STRAIN=CV. TOPAS; TISSUE=Pollen;
RX   MEDLINE=98345939; PubMed=9680961;
RA   Murphy D.J., Ross J.H.;
RT   "Biosynthesis, targeting and processing of oleosin-like proteins,
RT   which are major pollen coat components in Brassica napus.";
RL   Plant J. 13:1-16(1998).
CC   -!- FUNCTION: MAJOR COMPONENT OF THE POLLEN COAT.
CC   -!- TISSUE SPECIFICITY: POLLEN.
FT   NON_TER      15     15
SQ   SEQUENCE   15 AA;  1756 MW;  4C9B7C58ED18A442 CRC64;


  Query Match             26.5%;  Score 22.5;  DB 1;  Length 15;
  Best Local Similarity   55.6%;  Pred. No. 5.1e+02;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    1;  Gaps    1;

Qy       8 PLVMNEFMN 16
           ||  ||::|
Db       2 PLT-NEYLN 9


RESULT   3
TM2A_METMA
ID   TM2A_METMA     STANDARD;      PRT;    12 AA.
AC   P80652;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   ALTERNATIVE TETRAHYDROMETHANOPTERIN S-METHYLTRANSFERASE 28 KDA SUBUNIT
DE   (EC 2.1.1.86) (N5-METHYLTETRAHYDROMETHANOPTERIN--COENZYME M
DE   METHYLTRANSFERASE 28 KDA SUBUNIT) (FRAGMENT).
OS   Methanosarcina mazei (Methanosarcina frisia).
OC   Archaea; Euryarchaeota; Methanosarcinales; Methanosarcinaceae;
OC   Methanosarcina.
OX   NCBI_TaxID=2209;
RN   [1]
RP   SEQUENCE.
RC   STRAIN=DSM 3647 / GOE1;
RX   MEDLINE=96370840; PubMed=8774736;
RA   Lienard T., Becher B., Marschall M., Bowien S., Gottschalk G.;
RT   "Sodium ion translocation by N5-methyltetrahydromethanopterin:
RT   coenzyme M methyltransferase from Methanosarcina mazei Go1
RT   reconstituted in ether lipid liposomes.";
RL   Eur. J. Biochem. 239:857-864(1996).
CC   -!- FUNCTION: THIS ENZYME COMPLEX CATALYZES AN INTERMEDIATE STEP IN
CC       METHANOGENESIS, THE FORMATION OF METHYL-COENZYME M AND
CC       TETRAHYDROMETHANOPTERIN FROM COENZYME M AND N5-METHYL-
CC       TETRAHYDROMETHANOPTERIN.
CC   -!- CATALYTIC ACTIVITY: 5-METHYL-5,6,7,8-TETRAHYDROMETHANOPTERIN +
CC       2-MERCAPTOETHANESULFONATE = 5,6,7,8-TETRAHYDROMETHANOPTERIN +
CC       2-(METHYLTHIO)ETHANESULFONATE.
CC   -!- SUBUNIT: COMPOSED OF SIX DIFFERENT SUBUNITS.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
KW   Transferase; Methyltransferase; Transmembrane; Methanogenesis.
FT   NON_TER      12     12
SQ   SEQUENCE   12 AA;  1321 MW;  6DE4A5766232D76B CRC64;


  Query Match             25.9%;  Score 22;  DB 1;  Length 12;
  Best Local Similarity   44.4%;  Pred. No. 4.9e+02;
  Matches    4;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       3 LVPMNPLVM 11
           | |: ||::
Db       4 LEPVLPLIV 12


RESULT   4
UC29_MAIZE
ID   UC29_MAIZE     STANDARD;      PRT;    15 AA.
AC   P80635;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   UNKNOWN PROTEIN FROM 2D-PAGE OF ETIOLATED COLEOPTILE (SPOT 45)
DE   (FRAGMENT).
OS   Zea mays (Maize).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC   Panicoideae; Andropogoneae; Zea.
OX   NCBI_TaxID=4577;
RN   [1]
RP   SEQUENCE.
RC   TISSUE=Coleoptile;
RA   Touzet P., Riccardi F., Morin C., Damerval C., Huet J.-C.,
RA   Pernollet J.-C., Zivy M., de Vienne D.;
RT   "The maize two dimensional gel protein database: towards an integrated
RT   genome analysis program.";
RL   Theor. Appl. Genet. 93:997-1005(1996).
CC   -!- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED PI OF THIS UNKNOWN
CC       PROTEIN IS: 4.9, ITS MW IS: 37.6 KDA.
DR   Maize-2DPAGE; P80635; COLEOPTILE.
DR   MaizeDB; 123960; -.
FT   NON_TER      1      1
FT   NON_TER     15     15
SQ   SEQUENCE   15 AA;  1679 MW;  3D53086B16018BC1 CRC64;


  Query Match             24.1%;  Score 20.5;  DB 1;  Length 15;
  Best Local Similarity   42.9%;  Pred. No. 1.1e+03;
  Matches    6;  Conservative    4;  Mismatches    3;  Indels    1;  Gaps    1;

Qy       2 NLVPMNPLVMNEFM 15
           | ||: |||  :::
Db       3 NPVPI-PLVDIDYL 15


RESULT   5
UPA7_HUMAN
ID   UPA7_HUMAN     STANDARD;      PRT;     9 AA.
AC   P30093;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   UNKNOWN PROTEIN FROM 2D-PAGE OF PLASMA (SPOT 18) (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE.
RC   TISSUE=Plasma;
RX   MEDLINE=93092937; PubMed=1459097;
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 6, 2002, 09:38:06 ; Search time 19.06 Seconds
(without alignments)
63.945 Million cell updates/sec

Title: US-09-897-042-18
Perfect score: 85
Sequence: 1 MNLVPMNPLVMNEFMN 16

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 2758

Minimum DB seq length: 0
Maximum DB seq length: 16

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	35	41.2	16	2	A45454	ankyrin-binding gl
2	28	32.9	15	2	S10891	ubiquitin thiolest
3	28	32.9	15	2	PT0093	ubiquitin thiolest
4	23	27.1	9	2	A60356	118K stomach cance
5	22	25.9	11	2	S09024	carboxylesterase (
6	21	24.7	11	2	A48973	glucoamylase A1 (E
7	21	24.7	12	2	S29859	gene p10 protein -
8	20	23.5	9	2	PT0080	60K Ca binding pro
9	20	23.5	12	2	PN0046	ATP synthase D cha
10	20	23.5	14	2	B61309	lutropin beta chai
11	20	23.5	14	2	F83754	hypothetical prote
12	20	23.5	15	2	PA0060	protein QF200037 -
13	20	23.5	15	2	B61457	alpha-glucosidase
14	20	23.5	16	2	A24099	crystal protein, 2
15	19	22.4	5	2	B37325	pap fimbrial regul
16	19	22.4	8	2	A39308	glycine reductase
17	19	22.4	8	2	H41978	calliFMRFamide 8 -
18	19	22.4	11	2	I41978	calliFMRFamide 9 -
19	19	22.4	12	2	S74196	3-hydroxy-3-methyl
20	19	22.4	14	2	S21247	H+-transporting AT
21	19	22.4	14	2	S23376	collagen alpha cha
22	19	22.4	14	2	G33160	H+-transporting AT
23	19	22.4	15	2	PQ0545	capsid protein VP1
24	19	22.4	15	2	S27248	pseudogermin - whe
25	19	22.4	15	2	PA0090	protein QF100022 -
26	19	22.4	15	2	PA0055	protein QF200007 -
27	19	22.4	15	2	PA0092	protein QF200011 -
28	19	22.4	15	2	A31435	adherence lectin h
29	19	22.4	15	2	S71306	heat shock protein
30	19	22.4	15	2	B59137	protein Pf1 - gold
31	19	22.4	16	2	I57530	gene c-fms protein
32	19	22.4	16	2	S11760	flgL protein - Cau
33	19	22.4	16	2	PH1302	Ig heavy chain DJ
34	19	22.4	16	2	S09084	proteasome chain 3
35	18.5	21.8	11	2	F60409	substance P-like p
36	18.5	21.8	13	2	A32453	phloroglucinol red
37	18.5	21.8	16	2	B24099	crystal protein, 7
38	18	21.2	5	2	A44692	fulicin - giant Af
39	18	21.2	8	4	I54017	granulocyte-colony
40	18	21.2	11	2	D60409	kassinin-like pept
41	18	21.2	11	2	PC2330	cycloinulooligosac
42	18	21.2	11	2	PH1375	T antigen variant
43	18	21.2	11	2	PC2254	cytochrome P450 3A
44	18	21.2	12	2	S11286	exo-alpha-sialidas
45	18	21.2	13	2	JZVHP1	crabrolin - Europe
46	18	21.2	13	2	D61491	seed protein ws-4
47	18	21.2	13	2	A40207	cell surface glyco
48	18	21.2	13	2	B61620	locustamyotropin I
49	18	21.2	13	2	E61458	Ig kappa chain V-I
50	18	21.2	13	2	I49637	deoxynucleotidyltr

ALIGNMENTS

RESULT 1
A45454
ankyrin-binding glycoprotein ABGP, 186K (N-terminal) - rat (fragment)
C;Species: Rattus norvegicus (Norway rat)
C;Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 07-Feb-1997
C;Accession: A45454
R;Davis, J.Q.; McLaughlin, T.; Bennett, V.
J. Cell Biol. 121, 121-133, 1993
A;Title: Ankyrin-binding proteins related to nervous system cell adhesion molecules:
A;Reference number: A45454; MUID:93209979
A;Accession: A45454
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-16 <DAV>
A;Experimental source: brain
A;Note: sequence extracted from NCBI backbone (NCBIP:128327)
C;Keywords: glycoprotein

Query Match 41.2%; Score 35; DB 2; Length 16;
Best Local Similarity 50.0%; Pred. No. 6.1;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

```
Qy      4 VPMNPLVMNE 13
            :||:| : ||
Db      3 IPMDPSIQNE 12
```

RESULT 2
S10891
ubiquitin thiolesterase (EC 3.1.2.15) PGP9.5 - human (fragment)
N;Alternate names: ubiquitin carboxyl-terminal hydrolase, neuron-specific
C;Species: Homo sapiens (man)
C;Date: 20-Feb-1995 #sequence_revision 29-May-1998 #text_change 11-Jan-2000
C;Accession: S10891
R;Day, I.N.M.; Hinks, L.J.; Thompson, R.J.
Biochem. J. 268, 521-524, 1990
A;Title: The structure of the human gene encoding protein gene product 9.5 (PGP9.5),
A;Reference number: S10891; MUID:90303237
A;Accession: S10891
A;Molecule type: DNA
A;Residues: 1-15 <DAY>
A;Cross-references: EMBL:X17377; NID:g35441; PIDN:CAA35249.1; PID:g296799
C;Genetics:
A;Introns: 11/3
C;Superfamily: human ubiquitin thiolesterase
C;Keywords: thiolester hydrolase

```
  Query Match             32.9%;  Score 28;  DB 2;  Length 15;
  Best Local Similarity   46.7%;  Pred. No. 89;
  Matches    7;  Conservative    3;  Mismatches    3;  Indels    2;  Gaps    1;

Qy       1 MNLVPM--NPLVMNE 13
           | | ||  || ::|:
Db       1 MQLKPMEINPEMLNK 15


RESULT   3
PT0093
ubiquitin thiolesterase (EC 3.1.2.15) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 18-Sep-1998
C;Accession: PT0093; PN0045
R;Kawakami, T.; Uchida, T.; Sakai, T.; Kamo, M.; Morimasa, T.; Tsugita, A.
Submitted to JIPID, July 1998
C;Description: Proteome analysis of mouse brain.
A;Reference number: PT0091
A;Accession: PT0093
A;Molecule type: protein
A;Residues: 1-15 <KAW>
A;Experimental source: brain, striatum
R;Kato, H.
Kawasaki Igakkaishi 22, 245-259, 1996
A;Title: Analysis of proteins isolated by two dimensional electrophoresis of mouse neuro
A;Reference number: PN0041
A;Accession: PN0045
A;Molecule type: protein
A;Residues: 1-15 <KAT>
A;Experimental source: neuroblastoma cell
C;Comment: The molecular mass is 28,000 and the pI is 5.0.
C;Keywords: brain; thiolester hydrolase


  Query Match             32.9%;  Score 28;  DB 2;  Length 15;
  Best Local Similarity   46.7%;  Pred. No. 89;
  Matches    7;  Conservative    3;  Mismatches    3;  Indels    2;  Gaps    1;

Qy       1 MNLVPM--NPLVMNE 13
           | | ||  || ::|:
Db       1 MQLKPMEINPEMLNK 15


RESULT   4
A60356
118K stomach cancer antigen - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 17-Mar-1999
C;Accession: A60356
R;Shiraishi, Y.
Int. J. Cancer 45, 783-787, 1990
A;Title: Western blotting analysis for malignant lymphoma and stomach cancer antigens fr
A;Reference number: A60356; MUID:90216080
A;Accession: A60356
A;Molecule type: protein
A;Residues: 1-9 <SHI>
C;Keywords: glycoprotein


  Query Match             27.1%;  Score 23;  DB 2;  Length 9;
  Best Local Similarity   28.6%;  Pred. No. 2.2e+05;
  Matches    2;  Conservative    4;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       4 VPMNPLV 10
           :|: |::
Db       1 IPLKPVI 7


RESULT   5
S09024
carboxylesterase (EC 3.1.1.1), microsomal - golden hamster (fragment)
C;Species: Mesocricetus auratus (golden hamster)
C;Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 28-Apr-1993
C;Accession: S09024
R;Hosokawa, M.; Maki, T.; Satoh, T.
Arch. Biochem. Biophys. 277, 219-227, 1990
A;Title: Characterization of molecular species of liver microsomal carboxylesterases
A;Reference number: S09021; MUID:90179180
A;Accession: S09024
A;Molecule type: protein
A;Residues: 1-11 <HOS>
C;Keywords: carboxylic ester hydrolase


  Query Match             25.9%;  Score 22;  DB 2;  Length 11;
  Best Local Similarity   50.0%;  Pred. No. 6.5e+02;
  Matches    4;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       5 PMNPLVMN 12
           |  | |:|
Db       2 PSXPXVVN 9


RESULT   6
A48973
glucoamylase A1 (EC 3.2.1.-) - Chalara paradoxa (fragment)
C;Species: Chalara paradoxa
C;Date: 19-Dec-1993 #sequence_revision 25-Apr-1997 #text_change 25-Apr-1997
C;Accession: A48973
R;Monma, M.; Kainuma, K.
Carbohydr. Res. 227, 385-388, 1992
A;Title: Heterogeneity of the glucoamylase components of the raw-starch-digesting amy
A;Reference number: A48973; MUID:92361881
A;Accession: A48973
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-11 <MON>
A;Note: sequence extracted from NCBI backbone (NCBIP:110946)
C;Keywords: glycosidase; hydrolase


  Query Match             24.7%;  Score 21;  DB 2;  Length 11;
  Best Local Similarity   50.0%;  Pred. No. 9.6e+02;
  Matches    3;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 MNEFMN 16
           ::||:|
Db       3 LDEFVN 8


RESULT   7
S29859
gene p10 protein - Choristoneura fumiferana nuclear polyhedrosis virus (fragment)
C;Species: Choristoneura fumiferana nuclear polyhedrosis virus, CfMNPV
C;Date: 19-Mar-1997 #sequence_revision 24-Mar-1999 #text_change 24-Mar-1999
C;Accession: S29859
R;Hill, J.E.; Kuzio, J.; Wilson, J.A.; MacKinnon, E.A.; Faulkner, P.
Biochim. Biophys. Acta 1172, 187-189, 1993
A;Title: Nucleotide sequence of the p74 gene of a baculovirus pathogenic to the spruc
A;Reference number: S29849; MUID:93176808
A;Accession: S29859
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-12 <HIL>
A;Experimental source: strain Ireland


  Query Match             24.7%;  Score 21;  DB 2;  Length 12;
  Best Local Similarity   80.0%;  Pred. No. 1.1e+03;
  Matches    4;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       5 PMNPL 9
           | |||
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: January 6, 2002, 13:23:02 ; Search time 78.06 Seconds
(without alignments)
147.968 Million cell updates/sec

Title: US-09-897-042-17
Perfect score: 51
Sequence: 1 atgaacttggtaccgatgaa.....tgaacgaattcatgaactaa 51

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 498480

Minimum DB seq length: 0
Maximum DB seq length: 51

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	18	35.3	41	4	US-09-030-156-21	Sequence 21, Appl
	2	18	35.3	48	4	US-09-358-972-150	Sequence 150, App
c	3	16	31.4	46	5	PCT-US95-11684-24	Sequence 24, Appl
	4	16	31.4	47	2	US-08-305-764C-26	Sequence 26, Appl
	5	15.6	30.6	25	4	US-09-286-682A-9	Sequence 9, Appli
	6	15.4	30.2	36	4	US-09-255-368-38	Sequence 38, Appl
	7	14.8	29.0	27	1	US-08-758-306-868	Sequence 868, App
c	8	14.8	29.0	28	2	US-08-666-493-18	Sequence 18, Appl
c	9	14.8	29.0	28	2	US-08-666-493-23	Sequence 23, Appl
c	10	14.8	29.0	31	1	US-08-485-284A-13	Sequence 13, Appl
c	11	14.8	29.0	32	3	US-08-997-803-10	Sequence 10, Appl
c	12	14.8	29.0	32	3	US-08-845-546-17	Sequence 17, Appl
c	13	14.8	29.0	46	2	US-08-465-380-127	Sequence 127, App
c	14	14.8	29.0	46	2	US-08-486-397-127	Sequence 127, App
c	15	14.8	29.0	46	2	US-08-486-399-127	Sequence 127, App
c	16	14.8	29.0	46	2	US-08-461-965-127	Sequence 127, App
c	17	14.8	29.0	46	2	US-08-634-641-127	Sequence 127, App
c	18	14.8	29.0	46	3	US-09-249-471-127	Sequence 127, App
c	19	14.8	29.0	46	3	US-09-249-472-127	Sequence 127, App
c	20	14.8	29.0	46	3	US-09-249-451-127	Sequence 127, App
c	21	14.8	29.0	46	3	US-08-809-455-127	Sequence 127, App
c	22	14.8	29.0	46	3	US-09-249-461-127	Sequence 127, App
c	23	14.8	29.0	46	3	US-09-249-448-127	Sequence 127, App
	24	14.8	29.0	51	3	US-08-930-798-5	Sequence 5, Appli
	25	14.8	29.0	51	4	US-09-502-778A-5	Sequence 5, Appli
c	26	14.6	28.6	40	1	US-08-032-846-35	Sequence 35, Appl
	27	14.6	28.6	43	1	US-07-995-950-6	Sequence 6, Appli
	28	14.6	28.6	43	1	US-08-300-582-6	Sequence 6, Appli
	29	14.6	28.6	49	1	US-08-351-413-15	Sequence 15, Appl
	30	14.6	28.6	49	2	US-09-025-583-15	Sequence 15, Appl
c	31	14.4	28.2	37	1	US-08-244-378A-19	Sequence 19, Appl
c	32	14.4	28.2	38	2	US-08-782-760-4	Sequence 4, Appli
c	33	14.4	28.2	38	5	PCT-US96-00995-4	Sequence 4, Appli
	34	14.2	27.8	24	3	US-08-483-316-3	Sequence 3, Appli
	35	14.2	27.8	24	5	PCT-US95-12624-3	Sequence 3, Appli
	36	14.2	27.8	28	3	US-08-740-784-3	Sequence 3, Appli
c	37	14.2	27.8	36	2	US-08-669-721-1	Sequence 1, Appli
c	38	14.2	27.8	36	4	US-09-189-344-1	Sequence 1, Appli
	39	14.2	27.8	41	4	US-09-143-634-13	Sequence 13, Appl
	40	14.2	27.8	44	4	US-08-850-744-7	Sequence 7, Appli
c	41	14.2	27.8	44	4	US-08-850-744-8	Sequence 8, Appli
	42	14.2	27.8	45	6	5474933-5	Patent No. 5474933
	43	14.2	27.8	47	1	US-08-170-588-5	Sequence 5, Appli
	44	14.2	27.8	47	2	US-08-482-182-5	Sequence 5, Appli
c	45	14.2	27.8	47	2	US-08-482-182-73	Sequence 73, Appl
	46	14.2	27.8	50	2	US-08-821-559A-19	Sequence 19, Appl
c	47	14	27.5	30	3	US-08-777-708C-10	Sequence 10, Appl
c	48	14	27.5	36	2	US-09-018-576-10	Sequence 10, Appl
c	49	14	27.5	36	3	US-09-248-137-10	Sequence 10, Appl
c	50	14	27.5	41	4	US-09-030-156-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-09-030-156-21
; Sequence 21, Application US/09030156
; Patent No. 6207373
; GENERAL INFORMATION:
; APPLICANT: Sosnowski, Ronald G.
; TITLE OF INVENTION: METHODS AND APPARATUS FOR DETERMINATION OF LENGTH
; TITLE OF INVENTION: POLYMORPHISMS IN DNA
; FILE REFERENCE: DAVID B. MURPHY: Nanogen 221/157
; CURRENT APPLICATION NUMBER: US/09/030,156
; CURRENT FILING DATE: 1998-02-25
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 41
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Human /
; OTHER INFORMATION: Biotin
US-09-030-156-21

Query Match 35.3%; Score 18; DB 4; Length 41;
Best Local Similarity 70.6%; Pred. No. 1e+02;
Matches 24; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```
Qy      18 gaacccactagtaatgaacgaattcatgaactaa 51
           |||||| |    |||||| |||| |||||  ||
Db       6 gaaccctcactgaatgaatgaatgaatgaatgaa 39
```

RESULT 2
US-09-358-972-150
; Sequence 150, Application US/09358972
; Patent No. 6235480
; GENERAL INFORMATION:
; APPLICANT: Shultz, John W
; APPLICANT: Lewis, Martin K.
; APPLICANT: Lieppe, Donna
; APPLICANT: Mandrekar, Michelle
; APPLICANT: Kephart, Daniel
; APPLICANT: Rhodes, Richard B.
; APPLICANT: Andrews, Christine A.
; APPLICANT: Hartnett, James R.

```
;  APPLICANT: Gu, Trent
;  APPLICANT: Olson, Ryan J.
;  APPLICANT: Wood, Keith W.
;  APPLICANT: Welch, Roy
;  TITLE OF INVENTION: Nucleic Acid Detection
;  FILE REFERENCE: Pro-103 6868/75528
;  CURRENT APPLICATION NUMBER: US/09/358,972
;  CURRENT FILING DATE: 1999-07-22
;  EARLIER APPLICATION NUMBER: 09/252,436
;  EARLIER FILING DATE: 1999-02-18
;  EARLIER APPLICATION NUMBER: 09/042,287
;  EARLIER FILING DATE: 1998-03-13
;  NUMBER OF SEQ ID NOS: 290
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 150
;    LENGTH: 48
;    TYPE: DNA
;    ORGANISM: Artificial Sequence
;    FEATURE:
;    OTHER INFORMATION: Description of Artificial Sequence:TPOX allele
US-09-358-972-150


  Query Match             35.3%;  Score 18;  DB 4;  Length 48;
  Best Local Similarity   70.6%;  Pred. No. 1.1e+02;
  Matches   24;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      18 gaacccactagtaatgaacgaattcatgaactaa 51
           |||||| |    |||||| ||||  |||||  ||
Db      11 gaaccctcactgaatgaatgaatgaatgaatgaa 44


RESULT   3
PCT-US95-11684-24/c
; Sequence 24, Application PC/TUS9511684
;  GENERAL INFORMATION:
;    APPLICANT:  THE SCRIPPS RESEARCH INSTITUTE
;    TITLE OF INVENTION:  CYTOTACTIN DERIVATIVES THAT STIMULATE
;    TITLE OF INVENTION:  ATTACHMENT AND NEURITE OUTGROWTH, AND METHODS OF MAKING
;    TITLE OF INVENTION:  AND USING SAME
;    NUMBER OF SEQUENCES:  28
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  The Scripps Research Institute, Office of
;      ADDRESSEE:  Patent Counsel
;      STREET:  10666 North Torrey Pines Road, TPC 8
;      CITY:  La Jolla
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/11684
;      FILING DATE:  14-SEP-1995
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/308,359
;      FILING DATE:  16-SEP-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Logan, April C.
;      REGISTRATION NUMBER:  33,950
;      REFERENCE/DOCKET NUMBER:  BEC0019P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-554-2937
;      TELEFAX:  619-554-6312
;  INFORMATION FOR SEQ ID NO:  24:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  46 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
PCT-US95-11684-24


  Query Match             31.4%;  Score 16;  DB 5;  Length 46;
  Best Local Similarity   68.8%;  Pred. No. 6.4e+02;
  Matches   22;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      12 accgatgaacccactagtaatgaacgaattca 43
           ||| || ||  || ||| ||  |  |||||||
Db      38 ACCCATAAATTCAGTAGCAACCACAGAATTCA 7


RESULT   4
US-08-305-764C-26
; Sequence 26, Application US/08305764C
; Patent No. 5856090
;  GENERAL INFORMATION:
;    APPLICANT:  Epstein, David M.
;    TITLE OF INVENTION:  DNA METHYLASE LINKING REACTION
;    NUMBER OF SEQUENCES:  71
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THE SCRIPPS RESEARCH INSTITUTE
;      STREET:  10550 No. 5856090th Torrey Pines Road
;      CITY:  La Jolla
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/305,764C
;      FILING DATE:  09-SEP-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fitting, Thomas
;      REGISTRATION NUMBER:  34,163
;      REFERENCE/DOCKET NUMBER:  TSRI 440.0
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 784-2937
;      TELEFAX:  (619) 784-9399
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  47 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-305-764C-26


  Query Match             31.4%;  Score 16;  DB 2;  Length 47;
  Best Local Similarity   62.5%;  Pred. No. 6.4e+02;
  Matches   25;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy       9 ggtaccgatgaacccactagtaatgaacgaattcatgaac 48
           | | ||   |||  || |  ||||||| ||  |||| | |
Db       7 GATCCCCGGGAATTCAATGATAATGAATGACATCATTACC 46


RESULT   5
US-09-286-682A-9/c
; Sequence 9, Application US/09286682A
; Patent No. 6284464
; GENERAL INFORMATION:
;  APPLICANT:  Cao, Xu
;  APPLICANT:  Shi, Xingming
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: January 6, 2002, 14:05:27 ; Search time 143.66 Seconds
(without alignments)
304.355 Million cell updates/sec

Title:. US-09-897-042-17
Perfect score: 51
Sequence: 1 atgaacttggtaccgatgaa.....tgaacgaattcatgaactaa 51

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1000800

Minimum DB seq length: 0
Maximum DB seq length: 51

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : N_Geneseq_1101:*
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	19.6	38.4	31	15	AAQ65370	Fish gonadotropic
	2	18	35.3	48	21	AAA86939	TPOX allele oligon
c	3	17.8	34.9	49	22	AAC88849	Linker oligonucleo
	4	17.8	34.9	49	22	AAC88850	Linker oligonucleo
c	5	17.6	34.5	43	18	AAT49063	Rapamycin-dependen
c	6	17.4	34.1	41	21	AAZ54901	Neisseria species
	7	16	31.4	47	20	AAV55594	Primer for Sin I c
	8	16	31.4	47	21	AAZ67749	Human map-related
	9	16	31.4	51	22	AAH79721	Human DNA containi
	10	15.8	31.0	40	21	AAA63948	PCR primer for del
	11	15.8	31.0	41	19	AAV50855	Maize polymorphic
	12	15.8	31.0	46	20	AAX35823	Antisense oligonuc
	13	15.8	31.0	46	21	AAZ89729	P. horikoshi DNA p
c	14	15.6	30.6	25	13	AAQ34341	Upstream PCR prime
c	15	15.6	30.6	25	20	AAZ07959	Osteopontin Hoxc-8
	16	15.6	30.6	33	22	AAF84503	PCR primer used to
	17	15.6	30.6	47	21	AAZ69492	Human map-related
c	18	15.6	30.6	51	22	AAH37212	Human SNP flanking
	19	15.4	30.2	27	19	AAV97099	Potato citrate syn
c	20	15.4	30.2	36	21	AAZ94699	Neuropeptide FF re
	21	15.4	30.2	41	19	AAV47713	Maize polymorphic
c	22	15.4	30.2	44	21	AAZ48429	Primer specific fo
c	23	15.4	30.2	49	21	AAA59450	Oligonucleotide us
c	24	15.2	29.8	25	21	AAA68733	Bacteriophage 96 O
	25	15.2	29.8	47	21	AAZ67761	Human map-related
c	26	15	29.4	24	22	AAH48391	Clathrin 12 PCR pr
c	27	15	29.4	30	21	AAC83846	Collagen binding A
	28	15	29.4	32	20	AAX90651	Forward primer CTE
c	29	15	29.4	39	18	AAV06624	Transcription enha
	30	15	29.4	40	13	AAQ20906	Antisense oligonuc
	31	14.8	29.0	27	19	AAV94319	Mouse IL-2 recepto
	32	14.8	29.0	29	22	AAH93907	Human prostate-spe
c	33	14.8	29.0	31	14	AAQ48596	PCR primer for ade
c	34	14.8	29.0	32	19	AAV39168	Human GLP-2 recept
c	35	14.8	29.0	32	20	AAX88758	Human EDG-5 recept
	36	14.8	29.0	40	15	AAQ62481	Staphylococcus aur
	37	14.8	29.0	44	20	AAZ12478	PCR primer used to
c	38	14.8	29.0	46	17	AAT41490	Oligonucleotide pr
c	39	14.8	29.0	46	20	AAZ10490	PCR primer used to
c	40	14.8	29.0	46	21	AAA73630	Ancyclostoma canin
c	41	14.8	29.0	48	21	AAZ61563	Nucleotide sequenc
c	42	14.6	28.6	30	21	AAC95446	Cat flea HNC CLBP2
c	43	14.6	28.6	37	21	AAZ94249	Saccharomyces cere
	44	14.6	28.6	38	22	AAH96226	Human Chk1 ribozym
c	45	14.6	28.6	40	15	AAQ73561	PCR primer for hum
c	46	14.6	28.6	40	17	AAT43021	Human gamma interf
c	47	14.6	28.6	40	17	AAT43064	Human gamma interf
	48	14.6	28.6	40	22	AAF26047	Antibody YB-1 dire
c	49	14.6	28.6	44	18	AAT70891	Primer WD23 for ch
	50	14.6	28.6	44	22	AAF57071	P. furiosus cdc6 g

ALIGNMENTS

RESULT 1
AAQ65370/c
ID AAQ65370 standard; DNA; 31 BP.
XX
AC AAQ65370;
XX
DT 10-JAN-1995 (first entry)
XX
DE Fish gonadotropic hormone beta IC chain primer.
XX
KW Polymerase chain reaction; PCR; amplify; primer; alpha; beta I;
KW beta II; chain; fish; gonadotropin hormone; GTH; oviposition;
KW animal; promotion of maturation; ss.
XX
OS Synthetic.
XX
PN JP06107689-A.
XX
PD 19-APR-1994.
XX
PF 20-NOV-1991; 91JP-0354152.
XX
PR 20-NOV-1991; 91JP-0354152.
XX
PA (NIOC) NIPPON OIL CO LTD.
XX
DR WPI; 1994-163941/20.
XX
PT Fish gonadotropic hormone - and transformants producing it,

PT useful for maturation and oviposition promotion in fish
XX
PS Example 5; Page 6; 22pp; Japanese.
XX
CC The sequences given in AAQ65365-72 are primers which were used in the
CC amplification of alpha and beta chain coding region fragments of the
CC fish gonadotropin hormone (GTH) cDNA. GTH may be produced by
CC transforming a host organism with the amplified sequence and collecting
CC the protein from the culture medium. The GTH produced may be used for
CC the promotion of maturation and oviposition in fish and other animals.
CC See also AAQ65373-76.
XX
SQ Sequence 31 BP; 7 A; 6 C; 5 G; 13 T; 0 other;

Query Match 38.4%; Score 19.6; DB 15; Length 31;
Best Local Similarity 84.6%; Pred. No. 1.3e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 17 tgaacccactagtaatgaacgaattc 42
| |||||||||| |||| | ||||||
Db 28 TAAACCCACTAGAAATGTAAGAATTC 3

RESULT 2
AAA86939
ID AAA86939 standard; DNA; 48 BP.
XX
AC AAA86939;
XX
DT 15-JAN-2001 (first entry)
XX
DE TPOX allele oligonucleotide.
XX
KW Detection; nucleic acid hybrid; depolymerisation; analysis; SNP;
KW single nucleotide polymorphism; identification; viral load; probe;
KW genotyping; medical marker diagnostic; primer; target; mutation;
KW genetic disease; ss.
XX
OS Synthetic.
XX
PN WO200049180-A1.
XX
PD 24-AUG-2000.
XX
PF 18-FEB-2000; 2000WO-US04242.
XX
PR 18-FEB-1999; 99US-0252436.
PR 21-JUL-1999; 99US-0358972.
PR 25-AUG-1999; 99US-0383316.
XX
PA (PROM-) PROMEGA CORP.
XX
PI Shultz JW, Lewis MK, Leippe D, Mandrekar M, Kephart D, Rhodes RB;
PI Andrews CA, Hartnett JR, Gu T, Olson RJ, Wood KV, Welch R;
XX
DR WPI; 2000-565377/52.
XX
PT Determining presence or absence of a predetermined endogenous nucleic
PT acid sequence by using an enzyme that depolymerizes the 3' end of an
PT oligonucleotide probe hybridized to a target sequence to release
PT identifier nucleotides -
XX
PS Example; Page 338; 389pp; English.
XX
CC The present invention describes a method (M1) for determining the
CC presence or absence of a predetermined endogenous nucleic acid target
CC sequence (ENAT). The method comprises hybridising a probe having an
CC identifier nucleotide (IN) with ENAT which is treated with an enzyme
CC that depolymerises the 3' end of hybridised NA to release the INs.
CC M1 is used for determining the number of known sequence repeats present
CC in a nucleic acid target sequence in a nucleic acid sample. The method
CC is also useful for determining whether a nucleic acid target sequence in
CC a sample is an allele from a homozygous or heterozygous locus. The
CC method is also useful for detection of mutations, translocations and
CC SNPs in nucleic acids (including those associated with genetic disease),
CC determination of viral load, species identification, sample
CC contamination, and analysis of forensic samples. AAA86791 to AAA87079
CC and AAB12817 represent sequence which are used in the exemplification of
CC the present invention.
CC N.B. There is a discrepancy between the SEQ ID NO: and sequences given
CC in the examples, and the SEQ ID NO: and sequences given in the sequence
CC listing from the present invention.
XX
SQ Sequence 48 BP; 18 A; 7 C; 11 G; 12 T; 0 other;

Query Match 35.3%; Score 18; DB 21; Length 48;
Best Local Similarity 70.6%; Pred. No. 5.7e+02;
Matches 24; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 18 gaacccactagtaatgaacgaattcatgaactaa 51
|||||| | |||||| |||| ||||| ||
Db 11 gaaccctcactgaatgaatgaatgaatgaatgaa 44

RESULT 3
AAC88849/c
ID AAC88849 standard; DNA; 49 BP.
XX
AC AAC88849;
XX
DT 05-MAR-2001 (first entry)
XX
DE Linker oligonucleotide HindXba+.
XX
KW Adenovirus type 35; Ad35; adenovirus type 5; Ad5; gene delivery vehicle;
KW gene therapy; linker; ss.
XX
OS Synthetic.
XX
PN EP1054064-A1.
XX
PD 22-NOV-2000.
XX
PF 16-MAY-2000; 2000EP-0201738.
XX
PR 17-MAY-1999; 99EP-0201545.
XX
PA (INTR-) INTROGENE BV.
XX
PI Bout A, Vogels R, Havenga MJE;
XX
DR WPI; 2001-001097/01.
XX
PT Adenovirus derived gene delivery vehicles comprising at least one
PT element of adenovirus type 35, efficiently transfers genetic material
PT to a human cell without the inherent problem of pre-existing immunity -
XX
PS Example 7; Page 20; 138pp; English.
XX
CC The present sequence is a linker used in a gene delivery vehicle
CC comprising an element of adenovirus type 35 or a functional equivalent
CC of such an element. The element is responsible for avoiding or reducing
CC neutralising activity against adenoviral elements by the host to which
CC the gene is to be delivered. The vehicle can be used to deliver genes or
CC nucleic acids of interest to host cells. Use of the delivery system
CC efficiently transfers genetic material to a human cell without the
CC inherent problem of pre-existing immunity, found with previous serotypes.
XX
SQ Sequence 49 BP; 14 A; 13 C; 11 G; 11 T; 0 other;

Query Match 34.9%; Score 17.8; DB 22; Length 49;
Best Local Similarity 67.6%; Pred. No. 6.8e+02;
Matches 25; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: January 6, 2002, 13:18:52 ; Search time 1369.35 Seconds
(without alignments)
400.215 Million cell updates/sec

Title: US-09-897-042-17
Perfect score: 51
Sequence: 1 atgaacttggtaccgatgaa.....tgaacgaattcatgaactaa 51

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 83076

Minimum DB seq length: 0
Maximum DB seq length: 51

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_htc:*
10: gb_est1:*
11: gb_est2:*
12: gb_htc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	16.4	32.2	43	10	AA746752	AA746752 nx63e01.s
	2	15.8	31.0	28	13	AZ663964	AZ663964 1M0543G20
c	3	15.6	30.6	49	10	AJ237149	AJ237149 AJ237149
c	4	15.4	30.2	36	13	AZ427281	AZ427281 1M0209H10
c	5	15.4	30.2	43	13	AZ310736	AZ310736 1M0025O16
	6	15.2	29.8	40	13	AZ973766	AZ973766 2M0248C07
c	7	15.2	29.8	49	10	AA254733	AA254733 mz83g09.r
	8	15	29.4	43	10	AI868103	AI868103 to96a07.x
	9	14.8	29.0	29	13	AZ813429	AZ813429 2M0080O11
	10	14.8	29.0	32	13	AZ477577	AZ477577 1M0297F09
	11	14.6	28.6	37	10	AA954354	AA954354 on54g01.s
	12	14.6	28.6	39	13	AZ785720	AZ785720 2M0029G20
	13	14.6	28.6	39	13	AZ827624	AZ827624 2M0104J12
	14	14.6	28.6	51	13	AZ766562	AZ766562 1M0564A18
	15	14.4	28.2	25	13	AZ666829	AZ666829 1M0549C17
	16	14.4	28.2	45	13	AZ309922	AZ309922 1M0017C23
	17	14.4	28.2	46	10	AA500769	AA500769 vg01b03.r
	18	14.4	28.2	48	13	AZ377335	AZ377335 1M0131D09
c	19	14.4	28.2	49	10	AI327470	AI327470 ma80h03.y
c	20	14.2	27.8	49	11	BG151815	BG151815 nag64h05.
c	21	14.2	27.8	50	10	AU104069	AU104069 AU104069
c	22	14.2	27.8	50	10	AW643187	AW643187 cm27a03.w
c	23	14	27.5	40	10	BE217253	BE217253 258-T7 R.
	24	14	27.5	44	11	W25663	W25663 zc64e08.r1
	25	14	27.5	45	13	AZ772088	AZ772088 1M0574P09
	26	14	27.5	47	13	AZ974877	AZ974877 2M0249N10
c	27	14	27.5	50	10	AU102367	AU102367 AU102367
c	28	13.8	27.1	27	13	AZ598057	AZ598057 1M0412B09
	29	13.8	27.1	36	13	AZ650257	AZ650257 1M0520O18
	30	13.8	27.1	38	13	AZ800974	AZ800974 2M0059E16
c	31	13.8	27.1	42	13	AZ445143	AZ445143 1M0240E13
c	32	13.8	27.1	50	11	BG271426	BG271426 nai50f04.
c	33	13.6	26.7	24	13	AZ330762	AZ330762 1M0056L12
	34	13.6	26.7	39	13	AZ640839	AZ640839 1M0503L05
	35	13.6	26.7	41	13	AZ768121	AZ768121 1M0567N20
	36	13.6	26.7	43	11	D20671	D20671 HUMGS01647
c	37	13.6	26.7	51	11	BF507243	BF507243 5923P-1 P
c	38	13.6	26.7	51	13	AZ441063	AZ441063 1M0232E06
c	39	13.4	26.3	40	13	AZ439838	AZ439838 1M0230H07
c	40	13.4	26.3	44	13	AZ486826	AZ486826 1M0315I09
c	41	13.4	26.3	46	13	AZ498067	AZ498067 1M0335J12
c	42	13.4	26.3	50	10	AU106560	AU106560 AU106560
c	43	13.2	25.9	23	13	AZ339880	AZ339880 1M0071K01
c	44	13.2	25.9	23	13	AZ371086	AZ371086 1M0122D02
	45	13.2	25.9	32	13	AZ793379	AZ793379 2M0046P01
	46	13.2	25.9	34	10	AI583535	AI583535 ts12d01.x
c	47	13.2	25.9	36	13	AZ658185	AZ658185 1M0534M18
c	48	13.2	25.9	43	13	AZ607760	AZ607760 1M0430F22
	49	13.2	25.9	49	10	AA180620	AA180620 MBAFCG1B1
	50	13.2	25.9	50	11	BG152210	BG152210 nag74b02.

ALIGNMENTS

RESULT 1
AA746752
LOCUS AA746752 43 bp mRNA EST 27-JAN-1998
DEFINITION nx63e01.s1 NCI_CGAP_Alv1 Homo sapiens cDNA clone IMAGE:1266936 similar to SW:ATP6_HUMAN P00846 ATP SYNTHASE A CHAIN ;, mRNA sequence.
ACCESSION AA746752
VERSION AA746752.1 GI:2786738
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 43)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Lee Helman, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: David B. Krizman, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Trace considered overall poor quality

```
           Insert Length: 825   Std Error: 0.00
           Seq primer: -40m13 fwd. ET from Amersham
           High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .43
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1266936"
                     /clone_lib="NCI_CGAP_Alv1"
                     /tissue_type="alveolar rhabdomyosarcoma"
                     /lab_host="DH10B"
                     /note="Vector: pAMP10; mRNA made from alveolar
                     rhabdomyosarcoma, cDNA made by oligo-dT priming.
                     Non-directionally cloned. Size-selected on agarose gel,
                     average insert  size 600 bp. Reference:  Krizman et al.
                     (1996) Cancer Research 56:5380-5383."
BASE COUNT       13 a      17 c       5 g       8 t
ORIGIN
```

```
  Query Match             32.2%;  Score 16.4;  DB 10;  Length 43;
  Best Local Similarity   67.6%;  Pred. No. 6.1e+04;
  Matches   23;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      12 accgatgaacccactagtaatgaacgaattcatg 45
           ||||| || || ||||   || |||  |  ||||
Db       1 ACCGACGACCCAACTATCTATAAACCTAGCCATG 34
```

```
RESULT   2
AZ663964
LOCUS       AZ663964       28 bp    DNA             GSS       14-DEC-2000
DEFINITION  1M0543G20R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0543G20 R, DNA sequence.
ACCESSION   AZ663964
VERSION     AZ663964.1  GI:11801110
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 28)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0543  row: G  column: 20
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 28.
FEATURES             Location/Qualifiers
     source          1. .28
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0543G20"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT        9 a       4 c       5 g      10 t
ORIGIN
```

```
  Query Match             31.0%;  Score 15.8;  DB 13;  Length 28;
  Best Local Similarity   74.1%;  Pred. No. 9.6e+04;
  Matches   20;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       1 atgaacttggtaccgatgaacccacta 27
           |||||||||  | |  || || ||| ||
Db       2 ATGAACTTGTTGCTTATAAATCCAGTA 28
```

```
RESULT   3
AJ237149/c
LOCUS       AJ237149       49 bp    mRNA            EST       26-MAR-2001
DEFINITION  AJ237149 Emericella nidulans FGSC4 early sexual Emericella nidulans
            cDNA clone esd0510, mRNA sequence.
ACCESSION   AJ237149
VERSION     AJ237149.1  GI:4191118
KEYWORDS    EST.
SOURCE      Emericella nidulans.
  ORGANISM  Emericella nidulans
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
            Eurotiales; Trichocomaceae; Emericella.
REFERENCE   1  (bases 1 to 49)
  AUTHORS   Jeong,H., Han,D.M., Jahng,K. and Chae,K.
  TITLE     The rpl16a Gene for Ribosomal Protein L16A Identified from
            Expressed Sequence Tags Is Differentially Expressed during Sexual
            Development of Aspergillus nidulans
  JOURNAL   Fungal Genet. Biol. 31 (2), 69-78 (2000)
  MEDLINE   21110325
COMMENT     Contact: Chae KS
            Faculty of Biological Sciences
            Chonbuk National University
            Chonju 561-756, Republic of Korea
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .49
                     /organism="Emericella nidulans"
                     /strain="FGSC4"
                     /db_xref="taxon:162425"
                     /clone="esd0510"
                     /clone_lib="Emericella nidulans FGSC4 early sexual"
                     /dev_stage="early sexual"
BASE COUNT       15 a      13 c       8 g       9 t       4 others
ORIGIN
```

```
  Query Match             30.6%;  Score 15.6;  DB 10;  Length 49;
  Best Local Similarity   70.0%;  Pred. No. 1.2e+05;
  Matches   21;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       9 ggtaccgatgaacccactagtaatgaacga 38
           ||| |||||| |||| |  | | || | ||
Db      32 GGTTCCGATGGACCCGCATGCATTGTATGA 3
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: January 6, 2002, 13:20:22 ; Search time 1443.27 Seconds
(without alignments)
582.951 Million cell updates/sec

Title: US-09-897-042-17
Perfect score: 51
Sequence: 1 atgaacttggtaccgatgaa.....tgaacgaattcatgaactaa 51

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 562506

Minimum DB seq length: 0
Maximum DB seq length: 51

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vi:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rod:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rod:*
36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	Query Match	Length	DB	ID	Description
1		18	35.3	41	6	AR139408	AR139408 Sequence
2		18	35.3	48	6	AR153148	AR153148 Sequence
3	c	17.8	34.9	49	6	AX049948	AX049948 Sequence
4		17.8	34.9	49	6	AX049949	AX049949 Sequence
5	c	17.8	34.9	50	14	PGCR4G1	AF189255 Porcine g
6		17	33.3	39	6	I06895	I06895 Sequence 4
7	c	16.8	32.9	51	6	AX158274	AX158274 Sequence
8		16	31.4	47	6	AR026322	AR026322 Sequence
9		15.8	31.0	46	6	E32392	E32392 Thermostabl
10		15.6	30.6	33	6	AX113720	AX113720 Sequence
11	c	15.6	30.6	51	6	AX114885	AX114885 Sequence
12	c	15.4	30.2	51	6	AX157029	AX157029 Sequence
13	c	15.4	30.2	51	6	AX157031	AX157031 Sequence
14	c	15.4	30.2	51	6	AX157032	AX157032 Sequence
15		15.2	29.8	47	6	AX114386	AX114386 Sequence
16	c	15.2	29.8	51	6	AX158273	AX158273 Sequence
17		15.2	29.8	51	10	MUSBMP24M	D45014 Mouse gene
18	c	15	29.4	39	6	A66657	A66657 Sequence 2
19		14.8	29.0	27	6	AR040020	AR040020 Sequence
20	c	14.8	29.0	28	6	A50939	A50939 Sequence 18
21	c	14.8	29.0	28	6	A50944	A50944 Sequence 23
22	c	14.8	29.0	28	6	AR072776	AR072776 Sequence
23	c	14.8	29.0	28	6	AR072781	AR072781 Sequence
24	c	14.8	29.0	31	6	AR007177	AR007177 Sequence
25	c	14.8	29.0	32	6	AR099746	AR099746 Sequence
26		14.8	29.0	40	6	E07012	E07012 Primer. 9/1
27	c	14.8	29.0	46	6	AR031892	AR031892 Sequence
28	c	14.8	29.0	46	6	AR031948	AR031948 Sequence
29	c	14.8	29.0	46	6	AR036329	AR036329 Sequence
30	c	14.8	29.0	46	6	AR075014	AR075014 Sequence
31	c	14.8	29.0	46	6	AR103878	AR103878 Sequence
32		14.8	29.0	51	6	AR119105	AR119105 Sequence
33	c	14.6	28.6	34	9	S80814	S80814 gamma delta
34		14.6	28.6	40	6	AX069262	AX069262 Sequence
35	c	14.6	28.6	40	6	I63348	I63348 Sequence 35
36		14.6	28.6	43	6	A21364	A21364 Synthetic o
37		14.6	28.6	43	6	AR007507	AR007507 Sequence
38		14.6	28.6	44	6	AX081683	AX081683 Sequence
39		14.6	28.6	49	6	AR007525	AR007525 Sequence
40		14.6	28.6	49	6	AR084091	AR084091 Sequence
41		14.6	28.6	51	6	AX165586	AX165586 Sequence
42	c	14.4	28.2	37	6	AR009554	AR009554 Sequence
43	c	14.4	28.2	38	6	AR072910	AR072910 Sequence
44	c	14.4	28.2	48	6	AX062298	AX062298 Sequence
45	c	14.4	28.2	50	10	AF071704	AF071704 Mus muscu
46		14.4	28.2	51	6	AX115073	AX115073 Sequence
47		14.4	28.2	51	6	AX117017	AX117017 Sequence
48		14.2	27.8	28	6	AR097982	AR097982 Sequence
49	c	14.2	27.8	36	6	AR053365	AR053365 Sequence
50	c	14.2	27.8	36	6	AR131089	AR131089 Sequence

ALIGNMENTS

RESULT 1
AR139408
LOCUS AR139408 41 bp DNA PAT 16-JUN-2001
DEFINITION Sequence 21 from patent US 6207373.
ACCESSION AR139408
VERSION AR139408.1 GI:14481904
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 41)
AUTHORS Sosnowski,R.G. and Tu,E.
TITLE Methods for determining nature of repeat units in DNA
JOURNAL Patent: US 6207373-A 21 27-MAR-2001;
FEATURES Location/Qualifiers

```
     source          1. .41
                     /organism="unknown"
BASE COUNT       16 a      5 c     10 g     10 t
ORIGIN


  Query Match             35.3%;  Score 18;  DB 6;  Length 41;
  Best Local Similarity   70.6%;  Pred. No. 1.9e+04;
  Matches   24;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      18 gaacccactagtaatgaacgaattcatgaactaa 51
           |||||| |    |||||| ||||  |||||  ||
Db       6 GAACCCTCACTGAATGAATGAATGAATGAATGAA 39


RESULT   2
AR153148
LOCUS       AR153148       48 bp    DNA             PAT       08-AUG-2001
DEFINITION  Sequence 150 from patent US 6235480.
ACCESSION   AR153148
VERSION     AR153148.1  GI:15120680
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 48)
  AUTHORS   Shultz,J.William, Lewis,M.K., Leippe,D., Mandrekar,M., Kephart,D.,
            Rhodes,R.Byron, Andrews,C.Ann, Hartnett,J.Robert, Gu,T.,
            Olson,R.J., Wood,K.V. and Welch,R.
  TITLE     Detection of nucleic acid hybrids
  JOURNAL   Patent: US 6235480-A 150 22-MAY-2001;
FEATURES             Location/Qualifiers
     source          1. .48
                     /organism="unknown"
BASE COUNT       18 a      7 c     11 g     12 t
ORIGIN


  Query Match             35.3%;  Score 18;  DB 6;  Length 48;
  Best Local Similarity   70.6%;  Pred. No. 1.9e+04;
  Matches   24;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      18 gaacccactagtaatgaacgaattcatgaactaa 51
           |||||| |    |||||| ||||  |||||  ||
Db      11 GAACCCTCACTGAATGAATGAATGAATGAATGAA 44


RESULT   3
AX049948/c
LOCUS       AX049948       49 bp    DNA             PAT       12-JAN-2001
DEFINITION  Sequence 49 from Patent WO0070071.
ACCESSION   AX049948
VERSION     AX049948.1  GI:12226325
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 49)
  AUTHORS   Bout,A., Havenga,M.J. and Vogels,R.
  TITLE     Adenovirus derived gene delivery vehicles comprising at least one
            element of adenovirus type 35
  JOURNAL   Patent: WO 0070071-A 49 23-NOV-2000;
            Introgene B.V. (NL)
FEATURES             Location/Qualifiers
     source          1. .49
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="linker HindXba+"
BASE COUNT       14 a      13 c     11 g     11 t
ORIGIN


  Query Match             34.9%;  Score 17.8;  DB 6;  Length 49;
  Best Local Similarity   67.6%;  Pred. No. 2.2e+04;
  Matches   25;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy       2 tgaacttggtaccgatgaacccactagtaatgaacga 38
           | | ||||||||||| ||||  | ||||   | |  ||
Db      49 TAAGCTTGGTACCGGTGAATTCGCTAGCGTTAACGGA 13


RESULT   4
AX049949
LOCUS       AX049949       49 bp    DNA             PAT       12-JAN-2001
DEFINITION  Sequence 50 from Patent WO0070071.
ACCESSION   AX049949
VERSION     AX049949.1  GI:12226326
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 49)
  AUTHORS   Bout,A., Havenga,M.J. and Vogels,R.
  TITLE     Adenovirus derived gene delivery vehicles comprising at least one
            element of adenovirus type 35
  JOURNAL   Patent: WO 0070071-A 50 23-NOV-2000;
            Introgene B.V. (NL)
FEATURES             Location/Qualifiers
     source          1. .49
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="linker HindXba-"
BASE COUNT       11 a      11 c     13 g     14 t
ORIGIN


  Query Match             34.9%;  Score 17.8;  DB 6;  Length 49;
  Best Local Similarity   67.6%;  Pred. No. 2.2e+04;
  Matches   25;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy       2 tgaacttggtaccgatgaacccactagtaatgaacga 38
           | | ||||||||||| ||||  | ||||   | |  ||
Db       5 TAAGCTTGGTACCGGTGAATTCGCTAGCGTTAACGGA 41


RESULT   5
PGCR4G1/c
LOCUS       PGCR4G1        50 bp    RNA             VRL       28-JUN-2000
DEFINITION  Porcine group C rotavirus RNA4 segment, partial sequence.
ACCESSION   AF189255
VERSION     AF189255.1  GI:6273125
KEYWORDS    .
SEGMENT     1 of 2
SOURCE      Porcine group C rotavirus.
  ORGANISM  Porcine group C rotavirus
            Viruses; dsRNA viruses; Reoviridae; Rotavirus; Group C rotaviruses.
REFERENCE   1  (bases 1 to 50)
  AUTHORS   Jiang,B., Saif,L.J., Gentsch,J.R. and Glass,R.I.
  TITLE     Completion of the four large gene sequences of porcine group C
            Cowden rotavirus
  JOURNAL   Virus Genes 20 (2), 193-194 (2000)
  MEDLINE   20329284
   PUBMED   10872883
REFERENCE   2  (bases 1 to 50)
  AUTHORS   Jiang,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-SEP-1999) Viral Gastroenteritis Section, Centers for
            Disease Control & Prevention, 1600 Clifton Road NE, Atlanta, GA
            30333, USA
FEATURES             Location/Qualifiers
     source          1. .50
                     /organism="Porcine group C rotavirus"
                     /strain="Cowden"
                     /db_xref="taxon:10968"
                     /chromosome="RNA4"
                     /note="5' terminus"
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 6, 2002, 09:42:57 ; Search time 12.31 Seconds
(without alignments)
85.918 Million cell updates/sec

Title: US-09-897-042-16
Perfect score: 262
Sequence: 1 MNLVPMNPLVMNGFCRYPSH.....RLLTKPERRLSWLLPPLSNN 47

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 148599

Minimum DB seq length: 0
Maximum DB seq length: 47

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	43	16.4	27	6	5187077-29	Patent No. 5187077
2	43	16.4	27	6	5427925-27	Patent No. 5427925
3	42	16.0	35	2	US-08-310-912A-66	Sequence 66, Appl
4	42	16.0	35	3	US-08-841-089-66	Sequence 66, Appl
5	42	16.0	35	4	US-09-301-085-66	Sequence 66, Appl
6	42	16.0	35	5	PCT-US95-04570-66	Sequence 66, Appl
7	42	16.0	35	5	PCT-US95-04589-66	Sequence 66, Appl
8	41.5	15.8	30	2	US-08-637-759B-492	Sequence 492, App
9	41.5	15.8	30	3	US-08-871-355A-492	Sequence 492, App
10	41	15.6	39	2	US-08-637-759B-144	Sequence 144, App
11	41	15.6	39	3	US-08-871-355A-144	Sequence 144, App
12	41	15.6	40	5	PCT-US95-04910-10	Sequence 10, Appl
13	40.5	15.5	33	1	US-08-240-049B-18	Sequence 18, Appl
14	40.5	15.5	33	1	US-08-259-148A-22	Sequence 22, Appl
15	40.5	15.5	33	1	US-08-484-054-22	Sequence 22, Appl
16	40.5	15.5	33	2	US-07-876-941A-22	Sequence 22, Appl
17	40.5	15.5	33	3	US-08-478-507-14	Sequence 14, Appl
18	40.5	15.5	33	3	US-08-478-507-17	Sequence 17, Appl
19	40.5	15.5	33	4	US-08-542-634-20	Sequence 20, Appl
20	40.5	15.5	33	4	US-09-128-275A-14	Sequence 14, Appl
21	40.5	15.5	33	4	US-09-128-275A-17	Sequence 17, Appl
22	40.5	15.5	33	4	US-08-477-292-20	Sequence 20, Appl
23	40.5	15.5	33	5	PCT-US95-13703-20	Sequence 20, Appl
24	40	15.3	17	4	US-09-025-769B-261	Sequence 261, App
25	38	14.5	47	3	US-09-100-600A-69	Sequence 69, Appl
26	37	14.1	23	1	US-08-268-251-55	Sequence 55, Appl
27	37	14.1	23	5	PCT-US93-01112-55	Sequence 55, Appl
28	36.5	13.9	25	2	US-08-726-306A-98	Sequence 98, Appl
29	36.5	13.9	26	4	US-08-974-549A-57	Sequence 57, Appl
30	36	13.7	27	3	US-08-819-177-10	Sequence 10, Appl
31	36	13.7	31	3	US-08-725-459B-78	Sequence 78, Appl
32	36	13.7	47	3	US-09-100-600A-67	Sequence 67, Appl
33	36	13.7	47	3	US-09-100-600A-68	Sequence 68, Appl
34	35	13.4	10	2	US-08-902-623-56	Sequence 56, Appl
35	35	13.4	12	2	US-08-310-912A-45	Sequence 45, Appl
36	35	13.4	12	3	US-08-841-089-45	Sequence 45, Appl
37	35	13.4	12	4	US-09-301-085-45	Sequence 45, Appl
38	35	13.4	12	5	PCT-US95-04570-45	Sequence 45, Appl
39	35	13.4	12	5	PCT-US95-04589-45	Sequence 45, Appl
40	35	13.4	14	1	US-08-077-252B-16	Sequence 16, Appl
41	35	13.4	14	1	US-07-977-696C-39	Sequence 39, Appl
42	35	13.4	14	1	US-08-129-930B-39	Sequence 39, Appl
43	35	13.4	14	4	US-09-002-753A-16	Sequence 16, Appl
44	35	13.4	14	4	US-08-134-346A-26	Sequence 26, Appl
45	35	13.4	15	1	US-08-116-733-13	Sequence 13, Appl
46	35	13.4	15	1	US-08-469-615-10	Sequence 10, Appl
47	35	13.4	15	1	US-08-466-763-10	Sequence 10, Appl
48	35	13.4	15	2	US-08-411-142A-10	Sequence 10, Appl
49	35	13.4	31	4	US-08-602-999A-56	Sequence 56, Appl
50	35	13.4	31	4	US-08-278-865-56	Sequence 56, Appl

ALIGNMENTS

```
RESULT   1
5187077-29
;Patent No. 5187077
;     APPLICANT: GEARING, DAVID P.;GOUGH, NICHOLAS M.;HILTON,
;DOUGLAS J.;KING, JULIE A.;METCALF, DONALD;NICE, EDOUARD C.
;NICOLA, NICOS A.;SIMPSON, RICHARD J.;WILLSON, TRACY A.
;     TITLE OF INVENTION: LEUKEMIA INHIBITORY FACTOR
;     NUMBER OF SEQUENCES: 41
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/294,514
;       FILING DATE: 09-DEC-1988
;SEQ ID NO:29:
;       LENGTH: 27
5187077-29


  Query Match             16.4%;  Score 43;  DB 6;  Length 27;
  Best Local Similarity   29.6%;  Pred. No. 28;
  Matches    8;  Conservative    6;  Mismatches   13;  Indels    0;  Gaps    0;

Qy       2 NLVPMNPLVMNGFCRYPSHWRPLEQIR 28
           | :|: |:       |:| |    : ||:
Db       1 NPLPITPVXATXAIRHPCHGNLMNQIK 27


RESULT   2
5427925-27
;Patent No. 5427925
;     APPLICANT: GEARING, DAVID P.;GOUGH, NICHOLAS M.;HILTON,
;DOUGLAS J.;KING, JULIE A.;METCALF, DONALD;NICE, EDOUARD C.
;NICOLA, NICOS A.;SIMPSON, RICHARD J.;WILLSON, TRACY A.
;     TITLE OF INVENTION: RECOMBNIANT METHOD FOR MAKING
;LEUKEMIA INHIBITOR FACTOR
;     NUMBER OF SEQUENCES: 38
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/58,979
;       FILING DATE: 06-MAY-1993
;       APPLICATION NUMBER:  948,614
;       FILING DATE: 22-SEP-1992
;       APPLICATION NUMBER:  667,159
;       FILING DATE: 11-MAR-1991
;SEQ ID NO:27:
;       LENGTH: 27
5427925-27
```

```
  Query Match             16.4%;  Score 43;  DB 6;  Length 27;
  Best Local Similarity   29.6%;  Pred. No. 28;
  Matches    8;  Conservative    6;  Mismatches   13;  Indels    0;  Gaps    0;

Qy      2 NLVPMNPLVMNGFCRYPSHWRPLEQIR 28
          | :|: |:       |:| |   : ||:
Db      1 NPLPITPVXATXAIRHPCHGNLMNQIK 27


RESULT   3
US-08-310-912A-66
; Sequence 66, Application US/08310912A
; Patent No. 5981730
;  GENERAL INFORMATION:
;    APPLICANT:  Ausubel, Frederick M.
;    APPLICANT:  Staskawicz, Brian J.
     APPLICANT:  Brent, Andrew F.
     APPLICANT:  Dahlbeck, Douglas
     APPLICANT:  Katagiri, Fumiaki
;    APPLICANT:  Kunkel, Barbara N.
;    APPLICANT:  Mindrinos, Michael N.
;    APPLICANT:  Yu, Guo-Liang
;    TITLE OF INVENTION:  RPS2 GENE FAMILY, PRIMERS, PROBES, AND DETECTION
;    TITLE OF INVENTION:  METHODS
;    NUMBER OF SEQUENCES:  208
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2904
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/310,912A
;      FILING DATE:  September 22, 1994
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/227,360
;      FILING DATE:  April 13, 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lech, Karen F.
;      REGISTRATION NUMBER:  35,238
;      REFERENCE/DOCKET NUMBER:  00786/254001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 542-5070
;      TELEFAX:  (617) 542-8906
;      TELEX:  100254
;  INFORMATION FOR SEQ ID NO:  66:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  35 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-310-912A-66


  Query Match             16.0%;  Score 42;  DB 2;  Length 35;
  Best Local Similarity   50.0%;  Pred. No. 50;
  Matches    6;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      10 VMNGFCRYPSHW 21
           |  |||    ||
Db      24 VQEGFCHIEEHW 35


RESULT   4
```

```
US-08-841-089-66
; Sequence 66, Application US/08841089
; Patent No. 6127607
;  GENERAL INFORMATION:
;    APPLICANT:  Ausubel, Frederick M.
;    APPLICANT:  Staskawicz, Brian J.
;    APPLICANT:  Brent, Andrew F.
;    APPLICANT:  Dahlbeck, Douglas
;    APPLICANT:  Katagiri, Fumiaki
;    APPLICANT:  Kunkel, Barbara N.
;    APPLICANT:  Mindrinos, Michael N.
;    APPLICANT:  Yu, Guo-Liang
;    TITLE OF INVENTION:  RPS2 GENE AND USES THEREOF
;    NUMBER OF SEQUENCES:  106
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson
;      STREET:  225 Franklin Street Suite 3100
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2904
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/841,089
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/227,360
;      FILING DATE:  13-APR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clark, Paul T.
;      REGISTRATION NUMBER:  30,162
;      REFERENCE/DOCKET NUMBER:  00786/230001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 542-5070
;      TELEFAX:  (617) 542-8906
;      TELEX:  100254
;  INFORMATION FOR SEQ ID NO:  66:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  35 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-841-089-66


  Query Match             16.0%;  Score 42;  DB 3;  Length 35;
  Best Local Similarity   50.0%;  Pred. No. 50;
  Matches    6;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      10 VMNGFCRYPSHW 21
           |  |||    ||
Db      24 VQEGFCHIEEHW 35


RESULT   5
US-09-301-085-66
; Sequence 66, Application US/09301085
; Patent No. 6262248
; GENERAL INFORMATION:
;  APPLICANT: Ausubel, Frederick M.
;  APPLICANT: Staskawicz, Brian J.
;  APPLICANT: Brent, Andrew F.
;  APPLICANT: Dahlbeck, Douglas
;  APPLICANT: Katagiri, Fumiaki
;  APPLICANT: Kunkel, Barbara N.
;  APPLICANT: Mindrinos, Michael N.
;  APPLICANT: Yu, Guo-Liang
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 6, 2002, 09:41:07 ; Search time 25.18 Seconds
(without alignments)
138.262 Million cell updates/sec

Title: US-09-897-042-16
Perfect score: 262
Sequence: 1 MNLVPMNPLVMNGFCRYPSH.....RLLTKPERRLSWLLPPLSNN 47

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 275253

Minimum DB seq length: 0
Maximum DB seq length: 47

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : A_Geneseq_1101:*
1: /SIDS2/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	52	19.8	28	20	AAY27903	Human secreted pro
2	43.5	16.6	44	22	AAM13623	Peptide #57 encode
3	43.5	16.6	44	22	AAM26020	Peptide #57 encode
4	43.5	16.6	44	22	AAM01373	Peptide #55 encode
5	43	16.4	38	22	AAM21828	Peptide #8262 enco
6	43	16.4	38	22	AAM38152	Peptide #12189 enc
7	43	16.4	44	22	AAM17095	Peptide #3529 enco
8	43	16.4	44	22	AAM29585	Peptide #3622 enco
9	43	16.4	44	22	AAM04793	Peptide #3475 enco
10	42	16.0	42	22	AAM16945	Peptide #3379 enco
11	42	16.0	42	22	AAM29430	Peptide #3467 enco
12	42	16.0	42	22	AAM04638	Peptide #3320 enco
13	42	16.0	45	22	AAM35738	Peptide #9775 enco
14	41.5	15.8	45	21	AAB32477	Human secreted pro
15	41	15.6	39	20	AAY08250	Human cadherin-F p
16	41	15.6	39	22	AAM21211	Peptide #7645 enco
17	41	15.6	39	22	AAM37429	Peptide #11466 enc
18	41	15.6	40	16	AAR84654	STLVpan-p tax prot
19	40.5	15.5	33	17	AAR96100	Hepatitis E virus
20	40.5	15.5	33	18	AAW35825	Hepatitis E virus
21	40.5	15.5	33	19	AAW80199	ET-NANB (HEV) Mexi
22	40.5	15.5	33	19	AAW80201	ET-NANB (HEV) Mexi
23	40.5	15.5	33	19	AAW71212	Epitope of the Mex
24	40.5	15.5	33	19	AAW71214	Epitope of the Mex
25	40.5	15.5	33	20	AAY31387	HEV US-1 peptide e
26	40.5	15.5	33	21	AAB24122	Hepatitis E virus
27	40.5	15.5	33	21	AAB24124	Hepatitis E virus
28	40.5	15.5	33	22	AAB62525	HEV-Mexican strain
29	40.5	15.5	33	22	AAB62527	HEV-Mexican strain
30	40.5	15.5	38	22	AAM31144	Peptide #5181 enco
31	40.5	15.5	41	22	AAM37213	Peptide #11250 enc
32	40.5	15.5	44	20	AAY41334	Human secreted pro
33	40.5	15.5	47	22	AAM22104	Peptide #8538 enco
34	40.5	15.5	47	22	AAM38557	Peptide #12594 enc
35	40	15.3	27	22	AAM33360	Peptide #7397 enco
36	40	15.3	38	22	AAM15332	Peptide #1766 enco
37	40	15.3	38	22	AAM27807	Peptide #1844 enco
38	40	15.3	38	22	AAM03089	Peptide #1771 enco
39	39.5	15.1	31	18	AAW21999	Insulin C chain pe
40	39.5	15.1	37	22	AAM19436	Peptide #5870 enco
41	39.5	15.1	37	22	AAM32365	Peptide #6402 enco
42	39.5	15.1	39	22	AAM21726	Peptide #8160 enco
43	39.5	15.1	39	22	AAM38042	Peptide #12079 enc
44	39.5	15.1	40	22	AAM16694	Peptide #3128 enco
45	39.5	15.1	40	22	AAM29181	Peptide #3218 enco
46	39.5	15.1	40	22	AAM04412	Peptide #3094 enco
47	39	14.9	11	22	AAG68930	Human Chk2 kinase
48	39	14.9	11	22	AAG69154	Human Chk2 kinase
49	39	14.9	18	12	AAR10021	Peptide having gro
50	39	14.9	18	21	AAY65913	Transforming growt

ALIGNMENTS

RESULT 1
AAY27903
ID AAY27903 standard; Protein; 28 AA.
XX
AC AAY27903;
XX
DT 30-JUL-1999 (first entry)
XX
DE Human secreted protein encoded by gene No. 117.
XX
KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW developmental abnormality; foetal deficiency; blood; allergy; renal;
KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX
OS Homo sapiens.
XX
PN WO9924836-A1.
XX
PD 20-MAY-1999.
XX
PF 04-NOV-1998; 98WO-US23435.
XX
PR 17-NOV-1997; 97US-0066100.
PR 07-NOV-1997; 97US-0064900.

```
PR   07-NOV-1997;   97US-0064908.
PR   07-NOV-1997;   97US-0064911.
PR   07-NOV-1997;   97US-0064912.
PR   07-NOV-1997;   97US-0064983.
PR   07-NOV-1997;   97US-0064984.
PR   07-NOV-1997;   97US-0064985.
PR   07-NOV-1997;   97US-0064987.
PR   07-NOV-1997;   97US-0064988.
PR   17-NOV-1997;   97US-0066090.
PR   17-NOV-1997;   97US-0066094.
PR   17-NOV-1997;   97US-0066095.
PR   17-NOV-1997;   97US-0066089.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Carter KC,  Ebner R,  Endress GA,  Feng P,  Janat F;
PI   Kyaw H,  Lafleur DW,  Moore PA,  Ni J,  Olsen HS,  Rosen CA;
PI   Ruben SM,  Shi Y,  Soppet DR,  Wei Y;
XX
DR   WPI; 1999-337740/28.
DR   N-PSDB; AAX85049.
XX
PT   New human secreted proteins and coding sequences useful for treating
PT   disorders of the immune system and hyperproliferative disorders
XX
PS   Disclosure; Page 168; 507pp; English.
XX
CC   This sequence represents a secreted human protein encoded by the gene
CC   clone detailed in the descriptor line.  The gene can be used to generate
CC   fusion proteins by linking to the gene to a human immunoglobulin Fc
CC   portion (e.g. AAX84924) for increasing the stability of the fused
CC   protein as compared to the human protein only.
CC   The invention relates to 125 novel genes and their fragments (nucleic
CC   acid sequences: AAX84933-X85057; amino acid sequences AAY27567-Y27933)
CC   which are useful for preventing, treating or ameliorating  medical
CC   conditions e.g. by protein or gene therapy.  Also, pathological
CC   conditions can be diagnosed by determining the amount of the new
CC   polypeptides in a sample or by determining the presence of mutations in
CC   the new polynucleotides. Specific uses are described for each of the 125
CC   polynucleotides, based on which tissues they are most highly expressed in
CC   (see AAX84933 for described uses).
XX
SQ   Sequence   28 AA;


  Query Match             19.8%;  Score 52;  DB 20;  Length 28;
  Best Local Similarity   38.5%;  Pred. No. 6.4;
  Matches   10;  Conservative    3;  Mismatches    3;  Indels   10;  Gaps    2;

Qy       8 PLVMNGFCR----YPS------HWRP 23
           |: :||||     :|       ||||
Db       1 piplngfcesreffpdsgsvllhwrp 26


RESULT   2
AAM13623
ID   AAM13623 standard; Protein; 44 AA.
XX
AC   AAM13623;
XX
DT   12-OCT-2001  (first entry)
XX
DE   Peptide #57 encoded by probe for measuring cervical gene expression.
XX
KW   Probe; human; microarray; gene expression; cervical epithelial cell;
KW   cervical cancer.
XX
OS   Homo sapiens.
XX
PN   WO200157278-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00670.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488901/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human cervical epithelial cells -
XX
PS   Claim 27; SEQ ID No 18449; 487pp; English.
XX
CC   The present invention relates to human single exon nucleic acid probes
CC   (SENP: see AAI10068-AAI28459). The present sequence is a peptide encoded
CC   by one such probe. The SENPs are derived from human HeLa cells. The SENPs
CC   can be used to produce a single exon microarray, which can be used for
CC   measuring human gene expression in a sample derived from human cervical
CC   epithelial cells. By measuring gene expression, the probes are therefore
CC   useful in grading and/or staging of diseases of the cervix, notably
CC   cervical cancer.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   44 AA;


  Query Match             16.6%;  Score 43.5;  DB 22;  Length 44;
  Best Local Similarity   29.4%;  Pred. No. 1.3e+02;
  Matches   10;  Conservative    3;  Mismatches   10;  Indels   11;  Gaps    1;

Qy      18 PSHW----------RPLEQIRLLTKPERRLSWL 40
             |  |          ||  || :: :|  |  ||
Db       5 pgrwrhcarntgrrcrpwwqisvvprpssrpgwl 38


RESULT   3
AAM26020
ID   AAM26020 standard; Protein; 44 AA.
XX
AC   AAM26020;
XX
DT   17-OCT-2001  (first entry)
XX
DE   Peptide #57 encoded by probe for measuring placental gene expression.
XX
KW   Probe; microarray; human; placenta; antenatal diagnosis;
KW   genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200157272-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00663.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 6, 2002, 09:45:02 ; Search time 24.22 Seconds
(without alignments)
283.848 Million cell updates/sec

Title: US-09-897-042-16
Perfect score: 262
Sequence: 1 MNLVPMNPLVMNGFCRYPSH.....RLLTKPERRLSWLLPPLSNN 47

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 24961

Minimum DB seq length: 0
Maximum DB seq length: 47

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : SPTREMBL_17:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	47	17.9	45	5	Q9V8F9	Q9v8f9 drosophila
2	44.5	17.0	38	13	Q9PV48	Q9pv48 hippocampus
3	44.5	17.0	38	13	Q9PV42	Q9pv42 siganus sp.
4	44.5	17.0	38	13	Q9PV35	Q9pv35 ostracion s
5	44.5	17.0	39	13	Q9PV46	Q9pv46 dicentrarch
6	44.5	17.0	39	13	Q9PV45	Q9pv45 epinephelus
7	44.5	17.0	39	13	Q9PV37	Q9pv37 psettodes s
8	44.5	17.0	39	13	Q9PV34	Q9pv34 mola mola.
9	44	16.8	39	5	Q9V8G0	Q9v8g0 drosophila
10	43	16.4	40	2	Q47559	Q47559 escherichia
11	42	16.0	30	2	Q52621	Q52621 proteus vul
12	42	16.0	38	12	P90346	P90346 human t-cel
13	42	16.0	39	12	P89650	P89650 human t-cel
14	42	16.0	45	5	Q9V8F6	Q9v8f6 drosophila
15	42	16.0	46	5	O77150	O77150 drosophila
16	41	15.6	38	12	P89649	P89649 human t-cel
17	41	15.6	39	12	Q85092	Q85092 simian t-ce
18	41	15.6	40	5	Q9V8G2	Q9v8g2 drosophila
19	41	15.6	42	4	Q9BTP5	Q9btp5 homo sapien
20	41	15.6	47	6	O97761	O97761 elephantulu
21	40	15.3	24	12	Q9WJT0	Q9wjt0 bovine leuk
22	39.5	15.1	37	13	Q9PV51	Q9pv51 mugil japon
23	39.5	15.1	38	13	Q9PV49	Q9pv49 sargocentro
24	39.5	15.1	38	13	Q9PV44	Q9pv44 mullus sp.
25	39.5	15.1	38	13	Q9PV43	Q9pv43 dissostichu
26	39.5	15.1	38	13	Q9PV41	Q9pv41 acanthurus
27	39.5	15.1	38	13	Q9PV40	Q9pv40 thunnus sp.
28	39.5	15.1	40	13	Q9PV52	Q9pv52 lophius sp.
29	39	14.9	43	8	Q9GF82	Q9gf82 gnetum gnem
30	39	14.9	43	12	Q87340	Q87340 chimpanzee
31	39	14.9	43	12	Q87316	Q87316 chimpanzee
32	39	14.9	45	12	Q87348	Q87348 chimpanzee
33	39	14.9	46	12	Q87298	Q87298 chimpanzee
34	39	14.9	46	12	Q87409	Q87409 chimpanzee
35	39	14.9	47	12	Q87307	Q87307 chimpanzee
36	39	14.9	47	12	Q87319	Q87319 chimpanzee
37	39	14.9	47	12	Q87322	Q87322 chimpanzee
38	39	14.9	47	12	Q87334	Q87334 chimpanzee
39	39	14.9	47	12	Q87337	Q87337 chimpanzee
40	39	14.9	47	12	Q87355	Q87355 chimpanzee
41	39	14.9	47	12	Q87366	Q87366 chimpanzee
42	39	14.9	47	12	Q87370	Q87370 chimpanzee
43	39	14.9	47	12	Q87396	Q87396 chimpanzee
44	38	14.5	27	5	O46189	O46189 holopneuste
45	38	14.5	30	13	Q9PS06	Q9ps06 bothrops ja
46	38	14.5	38	12	P90459	P90459 simian t-ce
47	38	14.5	38	12	Q88391	Q88391 simian t-ce
48	38	14.5	38	12	Q88393	Q88393 simian t-ce
49	38	14.5	38	12	Q88395	Q88395 simian t-ce
50	38	14.5	38	12	Q88405	Q88405 simian t-ce

ALIGNMENTS

RESULT 1
Q9V8F9
ID Q9V8F9 PRELIMINARY; PRT; 45 AA.
AC Q9V8F9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE CG18106 PROTEIN.
GN IM2 OR CG18106.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

```
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003799; AAF57708.1; -.
DR   FlyBase; FBgn0025583; IM2.
SQ   SEQUENCE   45 AA;  4740 MW;  81986CD63B8135B7 CRC64;


  Query Match             17.9%;  Score 47;  DB 5;  Length 45;
  Best Local Similarity   45.0%;  Pred. No. 36;
  Matches    9;  Conservative    5;  Mismatches    2;  Indels    4;  Gaps    1;

Qy       2 NLVPMNP----LVMNGFCRY 17
           | ||::|    :|:|| |:|
Db      19 NAVPLSPDPGNVVINGDCKY 38


RESULT   2
Q9PV48
ID   Q9PV48      PRELIMINARY;      PRT;    38 AA.
AC   Q9PV48;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   DYSTROPHIN (FRAGMENT).
GN   DYST.
OS   Hippocampus sp.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Gasterosteiformes;
OC   Syngnathidae; Hippocampus.
OX   NCBI_TaxID=72047;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99398697; PubMed=10468597;
RA   Venkatesh B., Ning Y., Brenner S.;
RT   "Late changes in spliceosomal introns define clades in vertebrate
RT   evolution.";
RL   Proc. Natl. Acad. Sci. U.S.A. 96:10267-10271(1999).
DR   EMBL; AF137147; AAD53424.1; -.
FT   NON_TER       1      1
FT   NON_TER      38     38
SQ   SEQUENCE   38 AA;  4345 MW;  AADB3D41F62C0193 CRC64;


  Query Match             17.0%;  Score 44.5;  DB 13;  Length 38;
  Best Local Similarity   42.3%;  Pred. No. 65;
  Matches   11;  Conservative    2;  Mismatches    4;  Indels    9;  Gaps    1;

Qy      31 TKPERRLSWLL---------PPLSNN 47
           |  |  |:|||         ||:|||
Db       3 TALEEVLTWLLSAEDGLQSQPPISNN 28


RESULT   3
Q9PV42
ID   Q9PV42      PRELIMINARY;      PRT;    38 AA.
AC   Q9PV42;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   DYSTROPHIN (FRAGMENT).
GN   DYST.
OS   Siganus sp.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes;
OC   Acanthuroidei; Siganidae; Siganus.
OX   NCBI_TaxID=94227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99398697; PubMed=10468597;
RA   Venkatesh B., Ning Y., Brenner S.;
RT   "Late changes in spliceosomal introns define clades in vertebrate
RT   evolution.";
RL   Proc. Natl. Acad. Sci. U.S.A. 96:10267-10271(1999).
DR   EMBL; AF137153; AAD53430.1; -.
FT   NON_TER       1      1
FT   NON_TER      38     38
SQ   SEQUENCE   38 AA;  4315 MW;  AAC1E781F62C0193 CRC64;


  Query Match             17.0%;  Score 44.5;  DB 13;  Length 38;
  Best Local Similarity   42.3%;  Pred. No. 65;
  Matches   11;  Conservative    2;  Mismatches    4;  Indels    9;  Gaps    1;

Qy      31 TKPERRLSWLL---------PPLSNN 47
           |  |  |:|||         ||:|||
Db       3 TALEEVLTWLLSAEDGLQGQPPISNN 28


RESULT   4
Q9PV35
ID   Q9PV35      PRELIMINARY;      PRT;    38 AA.
AC   Q9PV35;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   DYSTROPHIN (FRAGMENT).
GN   DYST.
OS   Ostracion sp.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC   Ostraciidae; Ostracion.
OX   NCBI_TaxID=94239;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99398697; PubMed=10468597;
RA   Venkatesh B., Ning Y., Brenner S.;
RT   "Late changes in spliceosomal introns define clades in vertebrate
RT   evolution.";
RL   Proc. Natl. Acad. Sci. U.S.A. 96:10267-10271(1999).
DR   EMBL; AF137160; AAD53437.1; -.
FT   NON_TER       1      1
FT   NON_TER      38     38
SQ   SEQUENCE   38 AA;  4315 MW;  AAC1E781F62C0193 CRC64;


  Query Match             17.0%;  Score 44.5;  DB 13;  Length 38;
  Best Local Similarity   42.3%;  Pred. No. 65;
  Matches   11;  Conservative    2;  Mismatches    4;  Indels    9;  Gaps    1;

Qy      31 TKPERRLSWLL---------PPLSNN 47
           |  |  |:|||         ||:|||
Db       3 TALEEVLTWLLSAEDGLQGQPPISNN 28


RESULT   5
Q9PV46
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 6, 2002, 09:45:27 ; Search time 9.85 Seconds
(without alignments)
174.949 Million cell updates/sec

Title: US-09-897-042-16
Perfect score: 262
Sequence: 1 MNLVPMNPLVMNGFCRYPSH.....RLLTKPERRLSWLLPPLSNN 47

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 3287

imum DB seq length: 0
aximum DB seq length: 47

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID		Description
1	36	13.7	32	1	CRP_PLEPL	P12245	pleuronecte
2	35	13.4	24	1	CCAA_STRTI	P80436	streptomyce
3	35	13.4	41	1	CSP2_STROR	O33690	streptococc
4	34	13.0	43	1	BAGE_HUMAN	Q13072	homo sapien
5	34	13.0	43	1	PSBN_ZAMFU	Q9msr1	zamia furfu
6	34	13.0	46	1	YPC4_ECOLI	P19755	escherichia
7	33	12.6	23	1	TL17_SPIOL	P81778	spinacia ol
8	33	12.6	33	1	LPRH_ECOLI	P37324	escherichia
9	33	12.6	36	1	PSAI_CYACA	Q9tm24	cyanidium c
10	33	12.6	40	1	YSXC_SULAC	P39477	sulfolobus
11	32	12.2	42	1	NGF_VIPLE	P25428	vipera lebe
12	31.5	12.0	38	1	BD08_BOVIN	P46166	bos taurus
13	31.5	12.0	40	1	BD07_BOVIN	P46165	bos taurus
14	31	11.8	43	1	PSBN_ARATH	P12172	arabidopsis
15	31	11.8	43	1	PSBN_ORYSA	P12171	oryza sativ
16	31	11.8	43	1	PSBN_PEA	Q9xqr4	pisum sativ
17	30	11.5	25	1	ANT3_MESAU	P81050	mesocricetu
18	30	11.5	30	1	ITR2_ECBEL	P12071	ecballium e
19	30	11.5	36	1	PAHO_ALLMI	P06305	alligator m
20	30	11.5	36	1	PAHO_ANSAN	P06304	anser anser
21	30	11.5	36	1	PAHO_ERIEU	P41335	erinaceus e
22	30	11.5	43	1	PSBN_PINTH	P41626	pinus thunb
23	30	11.5	44	1	RS7_SALSA	P47839	salmo salar
24	29.5	11.3	45	1	YA78_ARCFU	O29185	archaeoglob
25	29	11.1	34	1	AMP2_CHICK	P80390	gallus gall
26	29	11.1	36	1	PAHO_EQUZE	P38000	equus zebra
27	29	11.1	39	1	GLL2_CHICK	P46157	gallus gall
28	29	11.1	47	1	HSP1_CAVPO	P35304	cavia porce
29	28.5	10.9	37	1	SCKA_TITSE	P46114	tityus serr
30	28.5	10.9	39	1	SCKM_CENMA	P40755	centruroide
31	28.5	10.9	43	1	CPC6_CANPG	P81585	cancer pagu
32	28.5	10.9	43	1	Y5K_CYMV	P16485	clover yell
33	28.5	10.9	46	1	HSP1_ORCOR	P24713	orcinus orc
34	28.5	10.9	47	1	H4Y_BLEJA	P80738	blepharisma
35	28	10.7	38	1	VG8_SPV4	P11340	spiroplasma
36	28	10.7	41	1	CSP_STROR	O33689	streptococc
37	28	10.7	42	1	DLP1_ORNAN	P82172	ornithorhyn
38	28	10.7	47	1	RK14_VIGUN	P26820	vigna ungui
39	27.5	10.5	35	1	IAAC_HORVU	P34951	hordeum vul
40	27.5	10.5	37	1	GHR3_RAT	P33580	rattus norv
41	27	10.3	15	1	UBL1_MONDO	P50103	monodelphis
42	27	10.3	28	1	ITR3_LUFCY	P35628	luffa cylin
43	27	10.3	28	1	TXO2_AGEAP	P15971	agelenopsis
44	27	10.3	32	1	RS20_THETH	P80380	thermus aqu
45	27	10.3	35	1	YQB5_CAEEL	Q09258	caenorhabdi
46	27	10.3	36	1	TX1B_AGEAP	P15970	agelenopsis
47	27	10.3	37	1	RL36_AQUAE	O66487	aquifex aeo
48	27	10.3	37	1	VG65_BPPH2	P16515	bacteriopha
49	27	10.3	37	1	YLZ8_CAEEL	P34421	caenorhabdi
50	27	10.3	39	1	DECO_MACDE	P17350	macrobdella

ALIGNMENTS

```
RESULT   1
CRP_PLEPL
ID   CRP_PLEPL      STANDARD;      PRT;    32 AA.
AC   P12245;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   01-OCT-1994 (Rel. 30, Last annotation update)
DE   C-REACTIVE PROTEIN (FRAGMENT).
OS   Pleuronectes platessa (Plaice).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
OC   Pleuronectoidei; Pleuronectidae; Pleuronectes.
OX   NCBI_TaxID=8262;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=82232106; PubMed=7093286;
RA   Pepys M.B., de Beer F.C., Milstein C.P., March J.F., Feinstein A.,
RA   Butress N., Clamp J.R., Taylor J., Bruton C., Fletcher T.C.;
RT   "C-reactive protein and serum amyloid P component in the plaice
RT   (Pleuronectes platessa L.), a marine teleost, are homologous with
RT   their human counterparts.";
RL   Biochim. Biophys. Acta 704:123-133(1982).
CC   -!- FUNCTION: CRP DISPLAYS SEVERAL FUNCTIONS ASSOCIATED WITH HOST
CC       DEFENSE: IT PROMOTES AGGLUTINATION, BACTERIAL CAPSULAR SWELLING,
CC       PHAGOCYTOSIS, & COMPLEMENT FIXATION THROUGH ITS CALCIUM-DEPENDENT
CC       BINDING TO PHOSPHORYLCHOLINE.
CC   -!- SUBUNIT: HOMOPENTAMER. PENTAXIN (OR PENTRAXIN) HAVE A DISCOID
CC       ARRANGEMENT OF 5 NONCOVALENTLY BOUND SUBUNITS.
CC   -!- SIMILARITY: BELONGS TO THE PENTAXIN FAMILY.
DR   PIR; A05291; A05291.
DR   InterPro; IPR001759; Pentaxin.
DR   PROSITE; PS00289; PENTAXIN; PARTIAL.
KW   Acute phase; Calcium; Pentaxin.
FT   DOMAIN        1    >32       PENTAXIN.
FT   NON_TER      32     32
SQ   SEQUENCE   32 AA;  3658 MW;  EDB2E0221D4FA227 CRC64;


  Query Match             13.7%;  Score 36;  DB 1;  Length 32;
  Best Local Similarity   42.9%;  Pred. No. 2.2e+02;
  Matches    6;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 MNLVPMNPLVMNGF 14
           : |:|| || :  |
Db      17 VELIPMKPLNLRAF 30


RESULT   2
CCAA_STRTI
ID   CCAA_STRTI     STANDARD;      PRT;    24 AA.
```

```
AC   P80436;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   CHINOXALIN-2-CARBOXYLIC ACID ACTIVATING ENZYME (FRAGMENT).
OS   Streptomyces triostinicus.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=45399;
RN   [1]
RP   SEQUENCE.
RA   Pahl A., Schlumbohm W., Keller U.;
RL   Submitted (MAR-1995) to the SWISS-PROT data bank.
CC   -!- FUNCTION: INVOLVED IN TRIOSTIN BIOSYNTHESIS.
KW   Antibiotic biosynthesis.
FT   NON_TER      24     24
SQ   SEQUENCE   24 AA;  2900 MW;  91C222B657CEB6D1 CRC64;


  Query Match             13.4%;  Score 35;  DB 1;  Length 24;
  Best Local Similarity   29.0%;  Pred. No. 2.2e+02;
  Matches    9;  Conservative    5;  Mismatches    9;  Indels    8;  Gaps    1;

Qy      10 VMNGFCRYPSHWRPLEQIRLLTKPERRLSWL 40
           :::||  :| |        |  :  ||  ||
Db       1 MLDGFVPWPDH--------LADEYRRRGIWL 23


RESULT   3
CSP2_STROR
ID   CSP2_STROR     STANDARD;      PRT;    41 AA.
AC   O33690;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   COMPETENCE STIMULATING PEPTIDE PRECURSOR (CSP).
GN   COMC.
OS   Streptococcus oralis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=1303;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DSM 20066;
RX   MEDLINE=98012953; PubMed=9352904;
RA   Haevarstein L.S., Hakenbeck R., Gaustad P.;
RT   "Natural competence in the genus Streptococcus: evidence that
RT   streptococci can change pherotype by interspecies recombinational
RT   exchanges.";
RL   J. Bacteriol. 179:6589-6594(1997).
CC   -!- FUNCTION: ACTS AS A PHEROMONE, INDUCES CELLS TO DEVELOP COMPETENCE
CC       FOR GENETIC TRANSFORMATION.
CC   -!- SUBCELLULAR LOCATION: SECRETED.
CC   -!- SIMILARITY: BELONGS TO THE COMC FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ000874; CAA04364.1; -.
KW   Pheromone; Competence.
FT   PROPEP        1     24       POTENTIAL.
FT   CHAIN        25     41       COMPETENCE STIMULATING PEPTIDE.
SQ   SEQUENCE   41 AA;  4961 MW;  B202A064BCD451AD CRC64;


  Query Match             13.4%;  Score 35;  DB 1;  Length 41;
  Best Local Similarity   50.0%;  Pred. No. 3.7e+02;
  Matches    8;  Conservative    1;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      21 WRPLEQIRLLTKPERR 36
           ||  | || |  | |:
Db      26 WRISETIRNLIFPRRK 41


RESULT   4
BAGE_HUMAN
ID   BAGE_HUMAN     STANDARD;      PRT;    43 AA.
AC   Q13072;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   B MELANOMA ANTIGEN (ANTIGEN MZ2-BA).
GN   BAGE.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Melanoma;
RX   MEDLINE=95202592; PubMed=7895173;
RA   Boel P., Wildmann C., Sensi M.L., Brasseur R., Renauld J.-C.,
RA   Coulie P., Boon T., van der Bruggen P.;
RT   "BAGE: a new gene encoding an antigen recognized on human melanomas
RT   by cytolytic T lymphocytes.";
RL   Immunity 2:167-175(1995).
CC   -!- FUNCTION: NOT KNOWN. ANTIGEN RECOGNIZED ON A MELANOMA BY
CC       AUTOLOGOUS CYTOLYTIC T LYMPHOCYTES.
CC   -!- TISSUE SPECIFICITY: NOT EXPRESSED IN NORMAL TISSUES, EXCEPT IN
CC       TESTIS. EXPRESSED WITH SIGNIFICANT PROPORTION IN MELANOMAS, BUT
CC       ALSO IN TUMORS OF VARIOUS HISTOLOGICAL ORIGINS, SUCH AS BLADDER
CC       CARCINOMAS, HEAD AND NECK SQUAMOUS CELL CARCINOMAS, LUNG AND
CC       BREAST CARCINOMAS. NOT EXPRESSED IN RENAL, COLORECTAL AND
CC       PROSTATIC CARCINOMAS, LEUKEMIAS AND LYMPHOMAS. MORE FREQUENTLY
CC       EXPRESSED IN METASTATIC MELANOMAS THAN IN PRIMARY MELANOMAS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U19180; AAC50123.1; -.
DR   MIM; 605167; -.
KW   Antigen.
SQ   SEQUENCE   43 AA;  4810 MW;  36F3BBCE4012F1BB CRC64;


  Query Match             13.0%;  Score 34;  DB 1;  Length 43;
  Best Local Similarity   40.0%;  Pred. No. 5.2e+02;
  Matches    8;  Conservative    3;  Mismatches    9;  Indels    0;  Gaps    0;

Qy      24 LEQIRLLTKPERRLSWLLPP 43
           | | ||: :    :|| | |
Db      14 LLQARLMKEESPVVSWRLEP 33


RESULT   5
PSBN_ZAMFU
ID   PSBN_ZAMFU     STANDARD;      PRT;    43 AA.
AC   Q9MSR1;
DT   20-AUG-2001 (Rel. 40, Created)
DT   20-AUG-2001 (Rel. 40, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   PHOTOSYSTEM II REACTION CENTER N PROTEIN.
GN   PSBN.
OS   Zamia furfuracea.
OG   Chloroplast.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 6, 2002, 09:43:22 ; Search time 18.55 Seconds
(without alignments)
193.003 Million cell updates/sec

Title: US-09-897-042-16
Perfect score: 262
Sequence: 1 MNLVPMNPLVMNGFCRYPSH.....RLLTKPERRLSWLLPPLSNN 47

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 10378

Minimum DB seq length: 0
Maximum DB seq length: 47

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	42	16.0	44	2	S29785	NADH oxidase (H2O2
2	41	15.6	39	2	S77904	tax protein - simi
3	40	15.3	29	2	S54340	diazepam binding i
4	39	14.9	31	2	A59321	protein Bmk 33-1 [
5	38	14.5	30	2	A53088	factor IX/factor X
6	36	13.7	32	2	A05291	C-reactive protein
7	36	13.7	43	2	E64740	hypothetical prote
8	35	13.4	40	2	B96843	protein TEL1S.1 [i
9	35	13.4	41	2	A49964	hypothetical prote
10	34.5	13.2	35	2	A61582	corticotropin / li
11	34.5	13.2	40	2	S19539	triacylglycerol li
12	34	13.0	26	2	A60044	DNA-binding protei
13	34	13.0	33	2	T39160	hypothetical prote
14	34	13.0	40	2	I64846	SP-2A gamma - huma
15	34	13.0	40	2	S17574	rRNA N-glycosidase
16	34	13.0	43	2	T00159	hypothetical prote
17	34	13.0	46	2	S12084	hypothetical prote
18	33.5	12.8	45	2	PC1218	homeobox protein F
19	33	12.6	30	2	A25497	peptide aPY - Amer
20	33	12.6	33	2	A05111	probable rho leade
21	33	12.6	33	2	D86064	rho operon leader
22	33	12.6	40	2	S21044	hypothetical prote
23	33	12.6	43	2	I39690	neural cell adhesi
24	33	12.6	46	2	A85605	hypothetical prote
25	33	12.6	46	2	B85656	hypothetical prote
26	32.5	12.4	38	2	E82858	hypothetical prote
27	32	12.2	36	2	H32502	T-cell receptor de
28	32	12.2	38	2	H39690	neural cell adhesi
29	32	12.2	45	2	T01784	spectrin beta Tand
30	32	12.2	46	2	A35437	nicotinic acetylch
31	31.5	12.0	32	2	C56649	streptomycin B-lik
32	31.5	12.0	36	2	H70251	hypothetical prote
33	31.5	12.0	40	2	G45495	beta-defensin-7 -
34	31.5	12.0	40	2	I45495	beta-defensin-9 -
35	31.5	12.0	43	2	T46409	hypothetical prote
36	31.5	12.0	45	2	H82526	hypothetical prote
37	31.5	12.0	46	2	D81077	hypothetical prote
38	31	11.8	17	2	PS0454	38K protein 3129 -
39	31	11.8	25	2	S52124	alpha-gliadin S11
40	31	11.8	26	2	D82344	IS1004 transposase
41	31	11.8	27	2	E44629	homeotic protein H
42	31	11.8	27	2	C44621	homeotic protein H
43	31	11.8	30	2	B53088	factor IX/factor X
44	31	11.8	34	2	I38214	protein-serine/thr
45	31	11.8	40	2	A34122	lectin-related pro
46	31	11.8	42	2	T07557	hypothetical prote
47	31	11.8	43	1	F2RZN	photosystem II pro
48	31	11.8	43	2	S05216	photosystem II pro
49	31	11.8	43	2	S14142	photosystem II pro
50	31	11.8	43	2	S12133	photosystem II pro

ALIGNMENTS

RESULT 1
S29785
NADH oxidase (H2O2-forming) (EC 1.6.-.-) - Thermus aquaticus
C;Species: Thermus aquaticus
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 17-Mar-1999
C;Accession: S29785
R;Park, H.J.; Kreutzer, R.; Reiser, C.O.A.; Sprinzl, M.
Eur. J. Biochem. 211, 909, 1993
A;Title: Molecular cloning and nucleotide sequence of the gene encoding a H(2)O(2)-fo
A;Reference number: S29785; MUID:93170329
A;Accession: S29785
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-44 <PAR>
A;Cross-references: EMBL:X60110
A;Note: the source is designated as Thermus thermophilus
C;Keywords: NAD; oxidoreductase

Query Match 16.0%; Score 42; DB 2; Length 44;
Best Local Similarity 61.5%; Pred. No. 90;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

```
Qy      17 YPSHWRPLEQIRL 29
           |||| ||||:: |
Db      30 YPSHRLPLERVVL 42
```

RESULT 2
S77904
tax protein - simian immunodeficiency virus SIVagm (fragment)
C;Species: simian immunodeficiency virus SIVagm
C;Date: 19-Mar-1997 #sequence_revision 18-Jul-1997 #text_change 20-Sep-1999
C;Accession: S77904
R;Vandamme, A.M.
submitted to the EMBL Data Library, May 1994
A;Reference number: S44288
A;Accession: S77904
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-39 <VAN>
A;Cross-references: EMBL:Z32851; NID:g483445; PIDN:CAA83684.1; PID:g483446
A;Experimental source: cell-line PP 1664
A;Note: the source is designated as primate T-cell lymphotropic virus
C;Superfamily: leukemia virus trans-activating transcription regulator

```
  Query Match             15.6%;  Score 41;  DB 2;  Length 39;
  Best Local Similarity   36.7%;  Pred. No. 1.1e+02;
  Matches   11;  Conservative    2;  Mismatches   13;  Indels    4;  Gaps    1;

Qy      10 VMNGFCRYPSHWRPLEQIRLLTKPERRLSW 39
           |  | |    |   |    | | || :|:|
Db      12 VSGGLCSTRLHRHAL----LATCPEHQLTW 37


RESULT   3
S54340
diazepam binding inhibitor 2 - human
C;Species: Homo sapiens (man)
C;Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 07-May-1999
C;Accession: S54340
R;Kolmer, M.; Rovio, A.; Alho, H.
Biochem. J. 306, 327-330, 1995
A;Title: The characterization of two diazepam binding inhibitor (DBI) transcripts in hum
A;Reference number: S54340; MUID:95194304
A;Accession: S54340
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-29 <KOL>


  Query Match             15.3%;  Score 40;  DB 2;  Length 29;
  Best Local Similarity   77.8%;  Pred. No. 1e+02;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      39 WLLPPLSNN 47
           ||||| | |
Db       6 WLLPPASAN 14


RESULT   4
A59321
protein Bmk 33-1 [imported] - Manchurian scorpion
C;Species: Mesobuthus martensis (Manchurian scorpion)
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000
C;Accession: A59321
R;Xu, C.
submitted to the Protein Sequence Database, August 2000
A;Reference number: A59321
A;Accession: A59321
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-31 <XUC>


  Query Match             14.9%;  Score 39;  DB 2;  Length 31;
  Best Local Similarity   63.6%;  Pred. No. 1.5e+02;
  Matches    7;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       5 PMNPLVMNGFC 15
           | ||   ||||
Db      18 PGNPPCRNGFC 28


RESULT   5
A53088
factor IX/factor X-binding anticoagulant protein B chain - jararaca (fragment)
C;Species: Bothrops jararaca (jararaca)
C;Date: 13-Sep-1994 #sequence_revision 18-Nov-1994 #text_change 15-Jun-1996
C;Accession: A53088
R;Sekiya, F.; Atoda, H.; Morita, T.
Biochemistry 32, 6892-6897, 1993
A;Title: Isolation and characterization of an anticoagulant protein homologous to botroc
A;Reference number: A53088; MUID:93326575
A;Accession: A53088
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-30 <SEK>
A;Experimental source: venom
A;Note: sequence extracted from NCBI backbone (NCBIP:135337)


  Query Match             14.5%;  Score 38;  DB 2;  Length 30;
  Best Local Similarity   42.1%;  Pred. No. 1.9e+02;
  Matches    8;  Conservative    3;  Mismatches    6;  Indels    2;  Gaps    1;

Qy      18 PSHWRPLEQ--IRLLTKPE 34
           || | | |    |: |:|:
Db       3 PSDWSPYEGHCYRVFTEPQ 21


RESULT   6
A05291
C-reactive protein - plaice (fragment)
C;Species: Pleuronectes platessa (plaice)
C;Date: 05-Jun-1987 #sequence_revision 05-Jun-1987 #text_change 22-Apr-1995
C;Accession: A05291
R;Pepys, M.B.; de Beer, F.C.; Milstein, C.P.; March, J.F.; Feinstein, A.; Butress, N.
Biochim. Biophys. Acta 704, 123-133, 1982
A;Reference number: A90642; MUID:82232106
A;Contents: tentative sequence
A;Accession: A05291
A;Molecule type: protein
A;Residues: 1-32 <PEP>
C;Superfamily: C-reactive protein
C;Keywords: acute phase; pentraxin; plasma


  Query Match             13.7%;  Score 36;  DB 2;  Length 32;
  Best Local Similarity   42.9%;  Pred. No. 3.7e+02;
  Matches    6;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 MNLVPMNPLVMNGF 14
           : |:|| || :  |
Db      17 VELIPMKPLNLRAF 30


RESULT   7
E64740
hypothetical protein b0165 - Escherichia coli
C;Species: Escherichia coli
C;Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 08-Oct-1999
C;Accession: E64740
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617
A;Accession: E64740
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-43 <BLAT>
A;Cross-references: GB:AE000126; GB:U00096; NID:g1786358; PIDN:AAC73276.1; PID:g17863
A;Experimental source: strain K-12, substrain MG1655


  Query Match             13.7%;  Score 36;  DB 2;  Length 43;
  Best Local Similarity   46.2%;  Pred. No. 5.1e+02;
  Matches    6;  Conservative    2;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       8 PLVMNGFCRYPSH 20
           | :  |:||  ||
Db      21 PPIFAGYCRDDSH 33


RESULT   8
B96843
protein TEL1S.1 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: B96843
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 6, 2002, 09:25:56 ; Search time 12.1 Seconds
(without alignments)
13.018 Million cell updates/sec

Title: US-09-897-042-14
Perfect score: 37
Sequence: 1 MNLVPMN 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 35098

Minimum DB seq length: 0
Maximum DB seq length: 7

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	21	56.8	6	2	US-08-910-484-7	Sequence 7, Appli
2	20	54.1	6	3	US-08-335-844A-67	Sequence 67, Appl
3	19	51.4	5	4	US-09-236-160-33	Sequence 33, Appl
4	18	48.6	5	6	5210075-68	Patent No. 5210075
5	18	48.6	6	2	US-08-458-109-40	Sequence 40, Appl
6	18	48.6	6	4	US-08-557-006C-13	Sequence 13, Appl
7	17	45.9	5	4	US-09-236-160-48	Sequence 48, Appl
8	17	45.9	6	1	US-08-325-509-46	Sequence 46, Appl
9	17	45.9	6	1	US-08-595-718A-1	Sequence 1, Appli
10	17	45.9	6	1	US-08-595-718A-2	Sequence 2, Appli
11	17	45.9	6	1	US-08-595-718A-3	Sequence 3, Appli
12	17	45.9	6	1	US-08-595-718A-6	Sequence 6, Appli
13	17	45.9	6	1	US-08-595-718A-7	Sequence 7, Appli
14	17	45.9	6	1	US-08-595-718A-8	Sequence 8, Appli
15	17	45.9	6	1	US-08-447-010-23	Sequence 23, Appl
16	17	45.9	6	4	US-09-091-814-65	Sequence 65, Appl
17	17	45.9	7	4	US-09-371-710-22	Sequence 22, Appl
18	17	45.9	7	4	US-09-230-603-8	Sequence 8, Appli
19	17	45.9	7	4	US-09-648-386-22	Sequence 22, Appl
20	16	43.2	4	2	US-08-429-964-41	Sequence 41, Appl
21	16	43.2	4	4	US-09-053-003-55	Sequence 55, Appl
22	16	43.2	4	5	PCT-US93-08062-41	Sequence 41, Appl
23	16	43.2	5	1	US-08-033-873-19	Sequence 19, Appl
24	16	43.2	5	1	US-08-331-383-32	Sequence 32, Appl
25	16	43.2	5	1	US-08-549-008-41	Sequence 41, Appl
26	16	43.2	5	1	US-08-447-010-16	Sequence 16, Appl
27	16	43.2	5	2	US-08-356-832-19	Sequence 19, Appl
28	16	43.2	5	3	US-08-802-981-4	Sequence 4, Appli
29	16	43.2	5	4	US-08-988-705-19	Sequence 19, Appl
30	16	43.2	5	4	US-09-236-160-27	Sequence 27, Appl
31	16	43.2	6	1	US-08-414-520A-1	Sequence 1, Appli
32	16	43.2	6	1	US-08-260-199A-19	Sequence 19, Appl
33	16	43.2	6	1	US-08-260-199A-27	Sequence 27, Appl
34	16	43.2	6	1	US-08-260-199A-29	Sequence 29, Appl
35	16	43.2	6	1	US-08-408-604A-45	Sequence 45, Appl
36	16	43.2	6	1	US-08-408-604A-46	Sequence 46, Appl
37	16	43.2	6	3	US-08-802-981-187	Sequence 187, App
38	16	43.2	6	5	PCT-US93-09626-45	Sequence 45, Appl
39	16	43.2	6	5	PCT-US93-09626-46	Sequence 46, Appl
40	16	43.2	7	2	US-08-770-301A-7	Sequence 7, Appli
41	16	43.2	7	2	US-08-630-645-21	Sequence 21, Appl
42	16	43.2	7	3	US-09-175-581-7	Sequence 7, Appli
43	16	43.2	7	5	PCT-US96-10220-21	Sequence 21, Appl
44	15	40.5	4	1	US-07-906-349A-16	Sequence 16, Appl
45	15	40.5	4	1	US-08-384-618-6	Sequence 6, Appli
46	15	40.5	4	1	US-08-102-757-23	Sequence 23, Appl
47	15	40.5	4	1	US-08-002-202-7	Sequence 7, Appli
48	15	40.5	4	1	US-08-167-035-16	Sequence 16, Appl
49	15	40.5	4	1	US-08-255-272-21	Sequence 21, Appl
50	15	40.5	4	1	US-08-208-887A-16	Sequence 16, Appl

ALIGNMENTS

```
RESULT 1
US-08-910-484-7
; Sequence 7, Application US/08910484
; Patent No. 5914244
;  GENERAL INFORMATION:
;    APPLICANT:  Coen, Donald M.
;    APPLICANT:  He, Zuwen
;    TITLE OF INVENTION:  UL97 FUSION PROTEINS AND METHODS OF USE
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows 95
;      SOFTWARE:  FastSEQ for Windows Version 2.0b
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/910,484
;      FILING DATE:  25-JUL-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/022,888
;      FILING DATE:  25-JUL-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Freeman, John W.
;      REGISTRATION NUMBER:  29,066
;      REFERENCE/DOCKET NUMBER:  00246/202001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-5070
;      TELEFAX:  617/542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-910-484-7


  Query Match 56.8%;  Score 21;  DB 2;  Length 6;
```

```
  Best Local Similarity   50.0%;  Pred. No. 1.6e+05;
  Matches    3;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 NLVPMN 7
           :: |||
Db       1 DITPMN 6


RESULT   2
US-08-335-844A-67
; Sequence 67, Application US/08335844A
; Patent No. 6066503
;  GENERAL INFORMATION:
;    APPLICANT:  GRAHAM, MARGARET
;    APPLICANT:  SMITH, TREVOR STANLEY
;    APPLICANT:  MUNN, EDWARD ALBERT
;    APPLICANT:  KNOX, DAVID PATRICK
     APPLICANT:  OLIVER, JOANNA JANE
     APPLICANT:  NEWTON, SUSAN ELIZABETH
     TITLE OF INVENTION:  RECOMBINANT DNA MOLECULES ENCODING
;    TITLE OF INVENTION:  AMINOPEPTIDASE ENZYMES AND THEIR USE IN THE PREPARATION OF
;    TITLE OF INVENTION:  VACCINES AGAINST HELMINTH INFECTIONS
;    NUMBER OF SEQUENCES:  73
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Rothwell, Figg, Ernst & Kurz
;      STREET:  Suite 701-E, 555 Thirteenth St., N.W
;      CITY:  Washington
;      STATE:  D. C.
;      COUNTRY:  U.S.A.
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/335,844A
;      FILING DATE:  09-JAN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB PCT/GB93/00943
;      FILING DATE:  06-MAY-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9209936
;      FILING DATE:  08-MAY-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  WALKER, Barbara W.
;      REGISTRATION NUMBER:  35,400
;      REFERENCE/DOCKET NUMBER:  1181-223A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)783-6040
;      TELEFAX:  (202)783-6031
;  INFORMATION FOR SEQ ID NO:  67:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-335-844A-67


  Query Match             54.1%;  Score 20;  DB 3;  Length 6;
  Best Local Similarity   50.0%;  Pred. No. 1.6e+05;
  Matches    3;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 MNLVPM 6
           :|| |:
Db       1 LNLTPI 6


RESULT   3
US-09-236-160-33
; Sequence 33, Application US/09236160A
; Patent No. 6307090
; GENERAL INFORMATION:
;  APPLICANT: BURKE Jr., Terrence R.
;  TITLE OF INVENTION: ACYLATED OLIGOPEPTIDE DERIVATIVES HAVING CELL SIGNAL
;  TITLE OF INVENTION: INHIBITING ACTIVITY
;  FILE REFERENCE: 175895
;  CURRENT APPLICATION NUMBER: US/09/236,160A
;  CURRENT FILING DATE: 1999-01-22
;  NUMBER OF SEQ ID NOS: 55
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 33
;   LENGTH: 5
;   TYPE: PRT
;   ORGANISM: mammalian
US-09-236-160-33


  Query Match             51.4%;  Score 19;  DB 4;  Length 5;
  Best Local Similarity   75.0%;  Pred. No. 1.6e+05;
  Matches    3;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 VPMN 7
           :|||
Db       1 MPMN 4


RESULT   4
5210075-68
;Patent No. 5210075
;    APPLICANT: SCHOLTZ, WOLFGANG;CHIANG, SHIU-LANG;NAGARAJAN,
;GOBI;LOBL, THOMAS J.
;    TITLE OF INVENTION: INTERLEUKIN 6 ANTAGONIST PEPTIDES
;    NUMBER OF SEQUENCES: 69
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,868
;      FILING DATE: 16-FEB-1990
;SEQ ID NO:68:
;      LENGTH: 5
5210075-68


  Query Match             48.6%;  Score 18;  DB 6;  Length 5;
  Best Local Similarity   75.0%;  Pred. No. 1.6e+05;
  Matches    3;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 NLVP 5
           |:|
Db       1 NLLP 4


RESULT   5
US-08-458-109-40
; Sequence 40, Application US/08458109
; Patent No. 5968791
;  GENERAL INFORMATION:
;    APPLICANT:  Calgene, Inc.
;    TITLE OF INVENTION:  PLANT LYSOPHOSPHATIDIC
;    TITLE OF INVENTION:  ACID ACYLTRANSFERASES
;    NUMBER OF SEQUENCES:  44
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Calgene, Inc.
;      STREET:  1920 Fifth Street
;      CITY:  Davis
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  95616
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 2.0 MB
;      COMPUTER:  Apple Macintosh
;      OPERATING SYSTEM:  Macintosh 7.1
;      SOFTWARE:  Microsoft Word 5.1(a)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/458,109
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 6, 2002, 09:18:51 ; Search time 25.55 Seconds
(without alignments)
20.294 Million cell updates/sec

Title: US-09-897-042-14
Perfect score: 37
Sequence: 1 MNLVPMN 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 47201

Minimum DB seq length: 0
Maximum DB seq length: 7

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : A_Geneseq_1101:*
1: /SIDS2/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	21	56.8	5	19	AAW35997	Human myocardium t
2	20	54.1	6	19	AAW75338	Hexapeptide #3 bin
3	20	54.1	6	19	AAW75270	Hexapeptide #3 bin
4	20	54.1	7	22	AAB47356	Cell lysis fractio
5	18	48.6	6	16	AAR87710	Peptide #9 in nove
6	18	48.6	6	16	AAR64314	Internal sequence
7	18	48.6	6	21	AAY69214	Peptide frament (r
8	18	48.6	7	12	AAR10518	Fusinus inhibitory
9	18	48.6	7	19	AAW81797	Seq ID 18 from DE1
10	18	48.6	7	20	AAY01780	Active fragment of
11	17	45.9	5	18	AAW22532	Peptide #6 having
12	17	45.9	6	14	AAR30840	Sequence of peptid
13	17	45.9	6	18	AAW22527	Peptide #1 having
14	17	45.9	6	18	AAW22528	Peptide #2 having
15	17	45.9	6	18	AAW22529	Peptide #3 having
16	17	45.9	6	18	AAW22535	Peptide #9 tested
17	17	45.9	6	18	AAW22536	Peptide #10 tested
18	17	45.9	6	18	AAW22534	Peptide #8 tested
19	17	45.9	6	19	AAW77228	Human MHC Class II
20	17	45.9	6	21	AAB03606	Aspergillus niger
21	17	45.9	6	21	AAB03608	Aspergillus niger
22	17	45.9	6	21	AAB03610	Aspergillus niger
23	17	45.9	7	18	AAW15795	Adherence factor p
24	17	45.9	7	19	AAW50220	Recombinant beta-g
25	17	45.9	7	22	AAB49427	Peptide #6 used to
26	16	43.2	4	12	AAR15772	Farnesyl-protein t
27	16	43.2	4	15	AAR49769	Farnesyltransferas
28	16	43.2	4	16	AAR77833	Farnesyl transfera
29	16	43.2	4	17	AAW04459	Farnesyl transfera
30	16	43.2	5	14	AAR30834	Sequence of peptid
31	16	43.2	5	19	AAW82095	Fluorogenic protea
32	16	43.2	5	19	AAW46560	Peptide containing
33	16	43.2	5	22	AAG73337	Alpha-1 antitrypsi
34	16	43.2	6	16	AAR75677	Streptokinase-bind
35	16	43.2	6	17	AAY07294	Peptide against Pl
36	16	43.2	6	17	AAR97583	MAb PK99H/PAK pept
37	16	43.2	6	17	AAR97591	MAb PK99H/PAK pept
38	16	43.2	6	17	AAR97593	MAb PK99H/PAK pept
39	16	43.2	6	19	AAW82029	Fluorogenic protea
40	16	43.2	6	19	AAW78565	SH2 domain binding
41	16	43.2	6	19	AAW78566	SH2 domain binding
42	16	43.2	6	20	AAY03909	Hepatitis C inhibi
43	16	43.2	6	20	AAW90096	Rat CART (62-102)
44	16	43.2	6	21	AAB24302	Prostate tumour as
45	16	43.2	6	21	AAB26204	Protein-tyrosine k
46	16	43.2	6	22	AAB92152	Signal transductio
47	16	43.2	6	22	AAB92153	Signal transductio
48	16	43.2	7	11	AAR03864	Hypotensive peptid
49	16	43.2	7	16	AAR75676	Streptokinase-bind
50	16	43.2	7	16	AAR84394	Hepatitis C virus

ALIGNMENTS

RESULT 1
AAW35997
ID AAW35997 standard; peptide; 5 AA.
XX
AC AAW35997;
XX
DT 05-MAR-1998 (first entry)
XX
DE Human myocardium troponin T epitope.
XX
KW Human; myocardium; troponin T; monoclonal antibody; epitope;
KW skeletal muscle; immunoassay.
XX
OS Homo sapiens.
XX
PN JP09271392-A.
XX
PD 21-OCT-1997.
XX
PF 09-APR-1996; 96JP-0086932.
XX
PR 09-APR-1996; 96JP-0086932.
XX
PA (TERU) TERUMO CORP.
XX
DR WPI; 1998-003027/01.
XX
PT Anti-human myocardium troponin T monoclonal antibody - useful in
PT immunoassay for distinguishing myocardial from skeletal muscle

PT troponin T
XX
PS Claim 1; Page 1; 6pp; Japanese.
XX
CC The present sequence represents a peptide epitope of human myocardium
CC troponin T, which is recognised by a new anti-human myocardium troponin
CC T monoclonal antibody. In an example from the specification, a hybridoma
CC producing anti-human myocardium troponin T IgG monoclonal antibody was
CC screened. An antibody used in human myocardium troponin T measuring
CC system was selected. The epitope of human myocardium troponin T-specific
CC antibody was analysed. An enzymatic immunoassay for troponin T
CC measurement was carried out. The measurements in a chronic renal disease
CC patient and a diabetes patient of normal heart function were examined.
CC The effect of co-presence of skeletal muscle troponin T was evaluated.
CC The anti-human myocardium troponin T monoclonal antibody is used in an
CC immunoassay for the determination of troponin T in a solution. The
CC method can determine troponin T exactly with no effect of skeletal
CC muscle troponin T.
XX
SQ Sequence 5 AA;

Query Match 56.8%; Score 21; DB 19; Length 5;
Best Local Similarity 100.0%; Pred. No. 4.3e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 NLVP 5
||||
Db 2 nlvp 5

RESULT 2
AAW75338
ID AAW75338 standard; peptide; 6 AA.
XX
AC AAW75338;
XX
DT 02-FEB-1999 (first entry)
XX
DE Hexapeptide #3 binds anti-Ad5 fibre head MAb 1D6.3.
XX
KW Cellular receptor; virus; immobilisation; monoclonal antibody; fibre;
KW viral surface protein; hexapeptide expression library; adenovirus;
KW major histocompatibility complex; MHC; fibronectin; gene therapy;
KW genetic disease; acquired immune deficiency syndrome; AIDS; cancer.
XX
OS Synthetic.
XX
PN FR2758821-A1.
XX
PD 31-JUL-1998.
XX
PF 30-JAN-1997; 97FR-0001005.
XX
PR 30-JAN-1997; 97FR-0001005.
PR 09-SEP-1997; 97FR-0011166.
XX
PA (CNRS) CNRS CENT NAT RECH SCI.
XX
PI Boulanger P, Hong SS, Karayan L;
XX
DR WPI; 1998-416458/36.
XX
PT Use of peptide(s) for facilitating or modulating attachment of
PT adenovirus to cells - useful for, e.g. treating or preventing
PT infection and improving uptake of gene therapy vectors
XX
PS Example 2; Fig 1; 48pp; French.
XX
CC The invention relates to methods for selecting and identifying a
CC cellular receptor for a virus, by immobilising, on a support, a
CC monoclonal antibody targeted to a viral surface protein that determines
CC attachment of the virus to the receptor. The immobilised antibody is
CC incubated with a hexapeptide expression library and peptides bound to the
CC immobilised antibody are eluted by competitive binding with recombinant
CC fragments of the viral surface protein. In a reverse method, the viral
CC surface protein is immobilised and incubated with the peptide library. In
CC this case, the bound peptides are eluted by competitive binding with the
CC monoclonal antibody. The hexapeptides AAW75336-W75345 represent
CC peptides isolated by biopanning the library with the immobilised
CC monoclonal antibody (MAb) 1D6.3 and eluted using an adenovirus type 5
CC fibre head protein (AAW75334).
CC The methods are used to identify peptides from MHC Class I and/or
CC type III fibronectin proteins that allow or facilitate attachment by
CC adenovirus (Ad) to host cells and/or entry into the cells, and to
CC identify ligands that modulate Ad infection mediated by these peptides,
CC e.g. to treat or prevent Ad infections or to facilitate infection by Ad
CC gene therapy vectors used to treat genetic diseases, acquired immune
CC deficiency syndrome or cancer.
XX
SQ Sequence 6 AA;

Query Match 54.1%; Score 20; DB 19; Length 6;
Best Local Similarity 60.0%; Pred. No. 4.3e+05;
Matches 3; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 3 LVPMN 7
|:| |
Db 1 lipfn 5

RESULT 3
AAW75270
ID AAW75270 standard; peptide; 6 AA.
XX
AC AAW75270;
XX
DT 02-FEB-1999 (first entry)
XX
DE Hexapeptide #3 binds anti-Ad5 fibre head MAb 1D6.3.
XX
KW Cellular receptor; virus; immobilisation; monoclonal antibody; fibre;
KW viral surface protein; hexapeptide expression library; adenovirus;
KW major histocompatibility complex; MHC; fibronectin; gene therapy;
KW genetic disease; acquired immune deficiency syndrome; AIDS; cancer.
XX
OS Synthetic.
XX
PN FR2758885-A1.
XX
PD 31-JUL-1998.
XX
PF 30-JAN-1997; 97FR-0001005.
XX
PR 30-JAN-1997; 97FR-0001005.
PR 10-JUL-1997; 97FR-0008796.
XX
PA (CNRS) CNRS CENT NAT RECH SCI.
XX
PI Boulanger P, Hong SS, Karayan L;
XX
DR WPI; 1998-416493/36.
XX
PT Selection and identification of cellular receptors for viruses -
PT used to control attachment and entry of adenovirus into cells, e.g.
PT for treating infection or in gene therapy
XX
PS Example 2; Fig 1; 43pp; French.
XX
CC The invention relates to methods for selecting and identifying a
CC cellular receptor for a virus, by immobilising, on a support, a
CC monoclonal antibody targeted to a viral surface protein that determines
CC attachment of the virus to the receptor. The immobilised antibody is
CC incubated with a hexapeptide expression library and peptides bound to the
CC immobilised antibody are eluted by competitive binding with recombinant

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 6, 2002, 09:35:21 ; Search time 24.53 Seconds
(without alignments)
41.741 Million cell updates/sec

Title: US-09-897-042-14
Perfect score: 37
Sequence: 1 MNLVPMN 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 65

Minimum DB seq length: 0
Maximum DB seq length: 7

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : SPTREMBL_17:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	13	35.1	7	2	O50556	O50556 actinobacil
2	13	35.1	7	13	O42564	O42564 fugu rubrip
3	12	32.4	5	13	P82099	P82099 litoria rub
4	12	32.4	6	13	P82096	P82096 litoria rub
5	10	27.0	7	8	O98866	O98866 spinacia ol
6	10	27.0	7	12	Q07624	Q07624 rous sarcom
7	10	27.0	7	13	P82065	P82065 litoria rub
8	9	24.3	7	2	O07354	O07354 synechococc
9	9	24.3	7	2	P72081	P72081 nocardia la
10	9	24.3	7	10	P93233	P93233 lycopersico
11	9	24.3	7	12	Q66205	Q66205 porcine tra
12	8	21.6	4	11	Q08433	Q08433 rattus norv
13	8	21.6	7	2	Q54248	Q54248 streptomyce
14	8	21.6	7	10	Q9C5B3	Q9c5b3 arabidopsis
15	8	21.6	7	11	Q63668	Q63668 rattus norv
16	8	21.6	7	12	Q66113	Q66113 cherry leaf
17	8	21.6	7	12	Q67113	Q67113 influenza a
18	7	18.9	7	2	O34028	O34028 pseudomonas
19	7	18.9	7	8	P92421	P92421 psathyrosta
20	7	18.9	7	8	P92385	P92385 hordeum mar
21	7	18.9	7	8	P92210	P92210 agropyron c
22	7	18.9	7	8	P92214	P92214 amblyopyrum
23	7	18.9	7	8	P92218	P92218 australopyr
24	7	18.9	7	8	P92221	P92221 bromus iner
25	7	18.9	7	8	P92226	P92226 crithopsis
26	7	18.9	7	8	P92372	P92372 haynaldia v
27	7	18.9	7	8	P92381	P92381 hordeum bra
28	7	18.9	7	8	P92387	P92387 henrardia p
29	7	18.9	7	8	P92390	P92390 heteranthel
30	7	18.9	7	8	P92393	P92393 hordeum vul
31	7	18.9	7	8	P92425	P92425 pseudoroegn
32	7	18.9	7	8	P92427	P92427 peridictyon
33	7	18.9	7	8	P92430	P92430 aegilops ta
34	7	18.9	7	8	P92442	P92442 taeniatheru
35	7	18.9	7	8	P92440	P92440 thinopyrum
36	7	18.9	7	8	P92403	P92403 lophopyrum
37	7	18.9	7	13	P82101	P82101 litoria rub
38	6	16.2	5	10	Q99007	Q99007 hordeum vul
39	6	16.2	7	2	Q47505	Q47505 escherichia
40	6	16.2	7	6	Q28742	Q28742 oryctolagus
41	6	16.2	7	8	O99182	O99182 gnatholebia
42	6	16.2	7	12	Q65578	Q65578 bovine herp
43	5	13.5	6	4	Q08720	Q08720 homo sapien
44	5	13.5	6	10	P82541	P82541 spinacia ol
45	5	13.5	7	2	Q47029	Q47029 enterobacte
46	5	13.5	7	11	Q63480	Q63480 rattus norv
47	5	13.5	7	12	Q9YQ10	Q9yq10 porcine tra
48	4	10.8	5	13	P82070	P82070 litoria rub
49	4	10.8	5	13	P82073	P82073 litoria rub
50	4	10.8	5	13	P82100	P82100 litoria rub

ALIGNMENTS

RESULT 1
O50556
ID O50556 PRELIMINARY; PRT; 7 AA.
AC O50556;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE GLYA (FRAGMENT).
GN GLYA.
OS Actinobacillus actinomycetemcomitans (Haemophilus
OS actinomycetemcomitans).
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Actinobacillus.
OX NCBI_TaxID=714;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33384;
RX MEDLINE=96355846; PubMed=8751884;
RA Kolodrubetz D., Spitznagel J. Jr., Wang B., Phillips L.H., Jacobs C.,
RA Kraig E.;
RT "cis Elements and trans factors are both important in strain-specific
RT regulation of the leukotoxin gene in Actinobacillus
RT actinomycetemcomitans.";
RL Infect. Immun. 64:3451-3460(1996).
DR EMBL; U51862; AAB88721.1; -.
FT NON_TER 1 1
SQ SEQUENCE 7 AA; 832 MW; 6DCB42D767340420 CRC64;

Query Match 35.1%; Score 13; DB 2; Length 7;
Best Local Similarity 40.0%; Pred. No. 4.7e+05;
Matches 2; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```
Qy      2 NLVPM 6
          | :|:
Db      1 NRLPV 5
```

```
RESULT   2
O42564
ID   O42564      PRELIMINARY;      PRT;     7 AA.
AC   O42564;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   TRUNCATED VOLTAGE-GATED SODIUM CHANNEL ALPHA SUBUNIT (FRAGMENT).
GN   SCN8A.
OS   Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC   Tetraodontidae; Takifugu.
OX   NCBI_TaxID=31033;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97442476; PubMed=9295353;
RA   Plummer N.W., McBurney M.W., Meisler M.H.;
RT   "Alternative splicing of the sodium channel SCN8A predicts a truncated
RT   two-domain protein in fetal brain and non-neuronal cells.";
RL   J. Biol. Chem. 272:24008-24015(1997).
DR   EMBL; U97673; AAB80916.1; -.
KW   Ionic channel.
FT   NON_TER       1      1
SQ   SEQUENCE   7 AA;  730 MW;  75B72EA2C73772A0 CRC64;


  Query Match             35.1%;  Score 13;  DB 13;  Length 7;
  Best Local Similarity   66.7%;  Pred. No. 4.7e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      4 VPM 6
          ||:
Db      1 VPL 3


RESULT   3
P82099
ID   P82099      PRELIMINARY;      PRT;     5 AA.
AC   P82099;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   ELECTRIN 3.
OS   Litoria rubella (Desert tree frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Neobatrachia; Bufonoidea; Hylidae;
OC   Litoria.
OX   NCBI_TaxID=104895;
RN   [1]
RP   SEQUENCE.
RC   TISSUE=SKIN SECRETION;
RA   Wabnitz P.A., Bowie J.H., Tyler M.J., Wallace J.C.;
RT   "Peptides from the skin glands of the Australian buzzing tree frog
RT   Litori electrica. Comparison with the skin peptides from Litoria
RT   rubella.";
RL   Aust. J. Chem. 52:0-0(1999).
KW   Amphibian skin; Amidation.
FT   MOD_RES       5      5       AMIDATION.
SQ   SEQUENCE   5 AA;  630 MW;  668761F2C9A00000 CRC64;


  Query Match             32.4%;  Score 12;  DB 13;  Length 5;
  Best Local Similarity   100.0%;  Pred. No. 4.7e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      5 PM 6
          ||
Db      4 PM 5


RESULT   4
P82096
ID   P82096      PRELIMINARY;      PRT;     6 AA.
AC   P82096;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   ELECTRIN 1.
OS   Litoria rubella (Desert tree frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Neobatrachia; Bufonoidea; Hylidae;
OC   Litoria.
OX   NCBI_TaxID=104895;
RN   [1]
RP   SEQUENCE.
RC   TISSUE=SKIN SECRETION;
RA   Wabnitz P.A., Bowie J.H., Tyler M.J., Wallace J.C.;
RT   "Peptides from the skin glands of the Australian buzzing tree frog
RT   Litori electrica. Comparison with the skin peptides from Litoria
RT   rubella.";
RL   Aust. J. Chem. 52:0-0(1999).
KW   Amphibian skin; Amidation.
FT   MOD_RES       6      6       AMIDATION.
SQ   SEQUENCE   6 AA;  792 MW;  6683704772C9A000 CRC64;


  Query Match             32.4%;  Score 12;  DB 13;  Length 6;
  Best Local Similarity   66.7%;  Pred. No. 4.7e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      4 VPM 6
          ||:
Db      2 VPI 4


RESULT   5
O98866
ID   O98866      PRELIMINARY;   PRT;    7 AA.
AC   O98866;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   CYTOCHROME B/F SUBUNIT IV (FRAGMENT).
OS   Spinacia oleracea (Spinach).
OG   Chloroplast.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Caryophyllidae; Caryophyllales; Chenopodiaceae; Spinacia.
OX   NCBI_TaxID=3562;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=86120353; PubMed=3003688;
RA   Sijben-Mueller G., Hallick R.B., Alt J., Westhoff P., Herrmann R.G.;
RT   "Spinach plastid genes coding for initiation factor IF-1, ribosomal
RT   protein S11 and RNA polymerase alpha-subunit.";
RL   Nucleic Acids Res. 14:1029-1044(1986).
DR   EMBL; X03496; CAA27215.1; -.
KW   Chloroplast.
FT   NON_TER       1      1
SQ   SEQUENCE   7 AA;  907 MW;  644729D77409C420 CRC64;


  Query Match             27.0%;  Score 10;  DB 8;  Length 7;
  Best Local Similarity   50.0%;  Pred. No. 4.7e+05;
  Matches    2;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 NLVP 5
          |  |
Db      1 NFRP 4


RESULT   6
Q07624
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 6, 2002, 09:35:46 ; Search time 9.82 Seconds
(without alignments)
26.136 Million cell updates/sec

Title: US-09-897-042-14
Perfect score: 37
Sequence: 1 MNLVPMN 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 74

Minimum DB seq length: 0
Maximum DB seq length: 7

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	13	35.1	7	1	CARP_MYTED	P10420 mytilus edu
2	13	35.1	7	1	MYOM_APLCA	P15513 aplysia cal
3	12	32.4	6	1	CIP1_MYTED	P13736 mytilus edu
4	12	32.4	6	1	CIP2_MYTED	P13737 mytilus edu
5	12	32.4	6	1	TMOF_SARBU	P41495 sarcophaga
6	10	27.0	6	1	VP19_HSV1K	P23210 herpes simp
7	10	27.0	7	1	FARB_CALVO	P41866 calliphora
8	10	27.0	7	1	UN06_PINPS	P81675 pinus pinas
9	9	24.3	4	1	RM01_YEAST	P36515 saccharomyc
10	9	24.3	7	1	CCF1_ENTFA	P20104 enterococcu
11	9	24.3	7	1	FAR4_PANRE	P41875 panagrellus
12	9	24.3	7	1	MNP1_LEPDE	P42984 leptinotars
13	9	24.3	7	1	MSCL_SALTY	P39446 salmonella
14	8	21.6	5	1	PRCT_PERAM	P01373 periplaneta
15	8	21.6	6	1	TRPI_PSEPU	P36414 pseudomonas
16	7	18.9	3	1	THYL_PIG	P01151 sus scrofa
17	7	18.9	4	1	DCML_PSECH	P19916 pseudomonas
18	7	18.9	4	1	TUFT_HUMAN	P01858 homo sapien
19	7	18.9	5	1	BIOB_SALTY	P12678 salmonella
20	7	18.9	5	1	BPP7_BOTIN	P30425 bothrops in
21	7	18.9	5	1	PAP2_PARMA	P81864 pardachirus
22	7	18.9	5	1	SUGA_ACHDO	P19991 acheta dome
23	7	18.9	6	1	OVM_LEPDE	P42985 leptinotars
24	7	18.9	7	1	ALL3_CARMA	P81806 carcinus ma
25	7	18.9	7	1	ALL4_CARMA	P81807 carcinus ma
26	7	18.9	7	1	ALL5_CARMA	P81808 carcinus ma
27	7	18.9	7	1	CHOX_ALCSP	P16101 alcaligenes
28	7	18.9	7	1	FAR1_HELTI	P41871 helisoma tr
29	7	18.9	7	1	FAR1_PROCL	P38499 procambarus
30	7	18.9	7	1	FAR2_PROCL	P38498 procambarus
31	7	18.9	7	1	GFRP_MOUSE	P99025 mus musculu
32	7	18.9	7	1	LANC_CARUI	P36960 carnobacter
33	7	18.9	7	1	UF04_MOUSE	P38642 mus musculu
34	7	18.9	7	1	WWA1_ACHFU	P35919 achatina fu
35	7	18.9	7	1	WWA2_ACHFU	P35920 achatina fu
36	7	18.9	7	1	WWA3_ACHFU	P35921 achatina fu
37	6	16.2	5	1	TPIS_CANFA	P54714 canis famil
38	6	16.2	5	1	TRM3_ECOLI	P13973 escherichia
39	6	16.2	6	1	ASP2_LACSN	P82655 lactobacill
40	6	16.2	6	1	FARP_MONEX	P41966 moniezia ex
41	6	16.2	6	1	UN06_CLOPA	P81351 clostridium
42	6	16.2	7	1	FAR1_ASCSU	P31889 ascaris suu
43	6	16.2	7	1	IGAO_DACDE	P06294 dactylium d
44	5	13.5	4	1	DCMS_PSECH	P19918 pseudomonas
45	5	13.5	4	1	FAR4_HIRME	P42563 hirudo medi
46	5	13.5	4	1	FMRF_MACNI	P01162 macrocallis
47	5	13.5	5	1	BIOA_CITFR	P13071 citrobacter
48	5	13.5	5	1	BIOA_SALTY	P12677 salmonella
49	5	13.5	5	1	BIOB_CITFR	P12997 citrobacter
50	5	13.5	6	1	ACPH_RABIT	P25154 oryctolagus

ALIGNMENTS

RESULT 1
CARP_MYTED
ID CARP_MYTED STANDARD; PRT; 7 AA.
AC P10420;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-MAR-1989 (Rel. 10, Last annotation update)
DE CATCH-RELAXING PEPTIDE (CARP).
OS Mytilus edulis (Blue mussel).
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorphia; Mytiloida;
OC Mytiloidea; Mytilidae; Mytilus.
OX NCBI_TaxID=6550;
RN [1]
RP SEQUENCE.
RX MEDLINE=88052022; PubMed=3676797;
RA Hirata T., Kubota I., Takabatake I., Kawahara A., Shimamoto N.,
RA Muneoka Y.;
RT "Catch-relaxing peptide isolated from Mytilus pedal ganglia.";
RL Brain Res. 422:374-376(1987).
CC -!- FUNCTION: THIS PEPTIDE EXHIBITS BOTH POTENTIATING (CONTRACTION)
CC AND INHIBITORY (RELAXATION) EFFECTS ON THE ANTERIOR BYSSUS
CC RETRACTOR MUSCLE.
DR PIR; A29342; ECMUCR.
KW Hormone; Amidation.
FT MOD_RES 7 7 AMIDATION.
SQ SEQUENCE 7 AA; 831 MW; 6734072687669DB0 CRC64;

Query Match 35.1%; Score 13; DB 1; Length 7;
Best Local Similarity 66.7%; Pred. No. 1e+05;
Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy      4 VPM 6
          :||
Db      2 MPM 4
```

RESULT 2
MYOM_APLCA
ID MYOM_APLCA STANDARD; PRT; 7 AA.
AC P15513;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE MYOMODULIN (NEURON B16 PEPTIDE).
OS Aplysia californica (California sea hare).
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Opisthobranchia; Anaspidea;
OC Aplysiidae; Aplysia.
OX NCBI_TaxID=6500;
RN [1]
RP SEQUENCE.

```
RC   TISSUE=Buccal muscle;
RX   MEDLINE=87261010; PubMed=3474664;
RA   Cropper E.C., Tenenbaum R., Kolks M.A.G., Kupfermann I., Weiss K.R.;
RT   "Myomodulin: a bioactive neuropeptide present in an identified
RT   cholinergic buccal motor neuron of Aplysia.";
RL   Proc. Natl. Acad. Sci. U.S.A. 84:5483-5486(1987).
CC   -!- FUNCTION: EXOGENOUS APPLICATION OF THE PEPTIDE POTENTIATES ARC
CC       MUSCLE CONTRACTIONS.
DR   PIR; A28340; A28340.
KW   Neuropeptide; Amidation.
FT   MOD_RES       7       7       AMIDATION.
SQ   SEQUENCE   7 AA;  847 MW;  6734072685B68700 CRC64;


  Query Match             35.1%;  Score 13;  DB 1;  Length 7;
  Best Local Similarity   66.7%;  Pred. No. 1e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 PMN 7
           ||:
Db       1 PMS 3


RESULT   3
CIP1_MYTED
ID   CIP1_MYTED     STANDARD;      PRT;     6 AA.
AC   P13736;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   CONTRACTION-INHIBITING PEPTIDE I (MIP I).
OS   Mytilus edulis (Blue mussel).
OC   Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorphia; Mytiloida;
OC   Mytiloidea; Mytilidae; Mytilus.
OX   NCBI_TaxID=6550;
RN   [1]
RP   SEQUENCE.
RC   TISSUE=Pedal ganglion;
RX   MEDLINE=88240357; PubMed=3377776;
RA   Hirata T., Kubota I., Iwasawa N., Takabatake I., Ikeda T., Muneoka Y.;
RT   "Structures and actions of Mytilus inhibitory peptides.";
RL   Biochem. Biophys. Res. Commun. 152:1376-1382(1988).
CC   -!- FUNCTION: INHIBITORY ACTION ON CONTRACTIONS IN SEVERAL MOLLUSCAN
CC       MUSCLES.
CC   -!- SIMILARITY: TO MIP II.
DR   PIR; A27696; A27696.
KW   Hormone; Amidation.
FT   MOD_RES       6       6       AMIDATION.
SQ   SEQUENCE   6 AA;  637 MW;  72C9C68775B81000 CRC64;


  Query Match             32.4%;  Score 12;  DB 1;  Length 6;
  Best Local Similarity   100.0%;  Pred. No. 1e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 PM 6
           ||
Db       3 PM 4


RESULT   4
CIP2_MYTED
ID   CIP2_MYTED     STANDARD;      PRT;     6 AA.
AC   P13737;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   CONTRACTION-INHIBITING PEPTIDE II (MIP II).
OS   Mytilus edulis (Blue mussel).
OC   Eu[redacted]; Metazoa; Mollusca; Bivalvia; Pteriomorphia; Mytiloida;
OC   M[redacted]a; Mytilidae; Mytilus.
OX   [redacted]
RN   [redacted]
RP   SEQUENCE.
RC   TISSUE=Pedal ganglion;
RX   MEDLINE=88240357; PubMed=3377776;
RA   Hirata T., Kubota I., Iwasawa N., Takabatake I., Ikeda T., Muneoka Y.;
RT   "Structures and actions of Mytilus inhibitory peptides.";
RL   Biochem. Biophys. Res. Commun. 152:1376-1382(1988).
CC   -!- FUNCTION: INHIBITORY ACTION ON CONTRACTIONS IN SEVERAL MOLLUSCAN
CC       MUSCLES.
CC   -!- SIMILARITY: TO MIP I.
DR   PIR; B27696; B27696.
KW   Hormone; Amidation.
FT   MOD_RES       6       6       AMIDATION.
SQ   SEQUENCE   6 AA;  621 MW;  72C9C6876DD81000 CRC64;


  Query Match             32.4%;  Score 12;  DB 1;  Length 6;
  Best Local Similarity   100.0%;  Pred. No. 1e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 PM 6
           ||
Db       3 PM 4


RESULT   5
TMOF_SARBU
ID   TMOF_SARBU     STANDARD;      PRT;     6 AA.
AC   P41495;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   TRYPSIN-MODULATING OOSTATIC FACTOR (TMOF).
OS   Sarcophaga bullata (Grey flesh fly) (Neobellieria bullata).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Oestroidea; Sarcophagidae; Sarcophaga.
OX   NCBI_TaxID=7385;
RN   [1]
RP   SEQUENCE, AND SYNTHESIS.
RC   TISSUE=Ovary;
RX   MEDLINE=94211930; PubMed=8159807;
RA   Bylemans D., Borovsky D., Hunt D.F., Shabanowitz J., Grauwels L.,
RA   de Loof A.;
RT   "Sequencing and characterization of trypsin modulating oostatic
RT   factor (TMOF) from the ovaries of the grey fleshfly, Neobellieria
RT   (Sarcophaga) bullata.";
RL   Regul. Pept. 50:61-72(1994).
CC   -!- FUNCTION: HAS AN OOSTATIC ACTIVITY. INHIBITS TRYPSIN BIOSYNTHESIS
CC       IN THE MIDGUT WHICH INDIRECTLY REDUCES THE VITELLOGENIN
CC       CONCENTRATION IN THE HEMOLYMPH RESULTING IN INHIBITION OF OOCYTE
CC       DEVELOPMENT.
CC   -!- DEVELOPMENTAL STAGE: SYNTHESIZED AND RELEASED FROM FOLLICULAR
CC       EPITHELIUM AFTER A BLOOD MEAL.
KW   Hormone.
SQ   SEQUENCE   6 AA;  695 MW;  61E72451B7642000 CRC64;


  Query Match             32.4%;  Score 12;  DB 1;  Length 6;
  Best Local Similarity   66.7%;  Pred. No. 1e+05;
  Matches    2;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       5 PMN 7
           | |
Db       2 PTN 4


RESULT   6
VP19_HSV1K
ID   VP19_HSV1K     STANDARD;      PRT;     6 AA.
AC   P23210;
DT   01-NOV-1991 (Rel. 20, Created)
DT   01-NOV-1991 (Rel. 20, Last sequence update)
DT   01-DEC-1992 (Rel. 24, Last annotation update)
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 6, 2002, 09:30:51 ; Search time 18.86 Seconds
(without alignments)
28.273 Million cell updates/sec

Title: US-09-897-042-14
Perfect score: 37
Sequence: 1 MNLVPMN 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 455

Minimum DB seq length: 0
Maximum DB seq length: 7

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	16	43.2	7	2	I48086	DNA topoisomerase
2	13	35.1	4	2	T46627	hypothetical prote
3	13	35.1	6	2	I65546	MHC H2-L antigen -
4	13	35.1	7	2	ECMUCR	catch-relaxing pep
5	13	35.1	7	2	A28340	myomodulin - Calif
6	12	32.4	6	2	A27696	contraction-inhibi
7	12	32.4	6	2	B27696	contraction-inhibi
8	11	29.7	4	2	A61300	22K superhelical D
9	11	29.7	4	2	S43014	hypothetical prote
10	11	29.7	4	2	I54357	schwannomin - mous
11	11	29.7	5	2	A60521	glycogen phosphory
12	11	29.7	5	2	B22565	R-phycoerythrin al
13	11	29.7	5	2	PQ0009	angiotensin-conver
14	11	29.7	7	2	PQ0663	membrane protein -
15	11	29.7	7	2	S17976	glucose isomerase
16	11	29.7	7	2	E61491	seed protein ws-5
17	11	29.7	7	2	S78024	ribosomal protein
18	10	27.0	6	2	I49424	cytotoxic T-lympho
19	10	27.0	7	2	S71867	glutathione transf
20	10	27.0	7	2	A61081	tryptophyllin, bas
21	10	27.0	7	2	PC2370	probable H+-transp
22	10	27.0	7	2	PN0649	alpha-dextrin endo
23	10	27.0	7	2	PN0150	omega-gliadine 1'
24	10	27.0	7	2	B44787	calliFMRFamide 11
25	10	27.0	7	2	C56793	platelet glycoprot
26	10	27.0	7	2	S58797	serine/threonine-s
27	10	27.0	7	2	S45648	Na+-transporting A
28	9	24.3	3	2	T13892	cytochrome-c oxida
29	9	24.3	4	2	A32039	tyrosine-melanocyt
30	9	24.3	4	2	S17255	ribosomal protein
31	9	24.3	5	2	E42364	flagellar protein
32	9	24.3	5	2	S11127	phosphoprotein, bo
33	9	24.3	6	4	A35039	hypothetical colla
34	9	24.3	6	4	S15596	orf 3 rara 5'-regi
35	9	24.3	7	2	A30812	sex pheromone cCF1
36	9	24.3	7	2	A11483	aspartate transami
37	9	24.3	7	2	I48105	dihydrofolate redu
38	8	21.6	3	2	PQ0010	angiotensin-conver
39	8	21.6	3	2	I78890	tyrosine protein k
40	8	21.6	4	2	B43848	cell surface adhes
41	8	21.6	4	2	A40135	branched-chain-ami
42	8	21.6	5	1	HOROHA	proctolin - Americ
43	8	21.6	5	2	JN0862	peptidyl-dipeptida
44	8	21.6	5	2	JN0860	peptidyl-dipeptida
45	8	21.6	5	2	C41225	copper resistance
46	8	21.6	5	2	E60274	major protein anti
47	8	21.6	5	2	B37988	acid proteinase li
48	8	21.6	5	2	A37114	hypoxanthine phosp
49	8	21.6	5	2	A60411	proctolin - Atlant
50	8	21.6	5	2	S53595	hypothetical prote

ALIGNMENTS

RESULT 1
I48086
DNA topoisomerase II alpha - Chinese hamster (fragment)
C;Species: Cricetulus griseus (Chinese hamster)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 05-Nov-1999
C;Accession: I48086
R;Ng, S.W.; Eder, J.P.; Schnipper, L.E.; Chan, V.T.W.
J. Biol. Chem. 270, 25850-25858, 1995
A;Title: Molecular cloning and characterization of the promoter for the Chinese hamst
A;Reference number: I48086; MUID:96029684
A;Accession: I48086
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-7 <RES>
A;Cross-references: EMBL:U34196; NID:g1041231; PIDN:AAC52315.1; PID:g1041232

Query Match 43.2%; Score 16; DB 2; Length 7;
Best Local Similarity 50.0%; Pred. No. 2.2e+05;
Matches 3; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MNLVPM 6
| | |:
Db 1 MELSPL 6

RESULT 2
T46627
hypothetical protein c4 - loblolly pine
C;Species: Pinus taeda (loblolly pine)
C;Date: 18-Feb-2000 #sequence_revision 18-Feb-2000 #text_change 18-Feb-2000
C;Accession: T46627
R;Chang, S.; Puryea, J.; Funkhouser, E.A.; Newton, R.J.; Cairney, J.
submitted to the EMBL Data Library, July 1995
A;Description: Cloning of a chitinase homolog which lacks chitin binding sites and is
A;Reference number: Z23105
A;Accession: T46627
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-4 <CHA>
A;Cross-references: EMBL:U31309; NID:g974285; PID:g974292
A;Experimental source: strain s6PT2xs6PT3; 8 month seedlings

Query Match 35.1%; Score 13; DB 2; Length 4;
Best Local Similarity 75.0%; Pred. No. 2.2e+05;
Matches 3; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy      1 MNLV 4
          | ||
Db      1 MKLV 4


RESULT   3
I65546
MHC H2-L antigen - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 05-Nov-1999
C;Accession: I65546
R;Kimura, A.; Israel, A.; Le Bail, O.; Kourilsky, P.
Cell 44, 261-272, 1986
A;Title: Detailed analysis of the mouse H-2Kb promoter: Enhancer-like sequences and thei
A;Reference number: I52778; MUID:86106202
A;Accession: I65546
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-6 <RES>
A;Cross-references: GB:M12483; NID:g199565; PIDN:AAA39663.1; PID:g554234

  Query Match             35.1%;  Score 13;  DB 2;  Length 6;
  Best Local Similarity   66.7%;  Pred. No. 2.2e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 LVP 5
          :||
Db      1 MVP 3


RESULT   4
ECMUCR
catch-relaxing peptide - blue mussel
N;Alternate names: CARP
C;Species: Mytilus edulis (blue mussel)
C;Date: 20-Jun-2000 #sequence_revision 20-Jun-2000 #text_change 20-Jun-2000
C;Accession: A29342
R;Hirata, T.; Kubota, I.; Takabatake, I.; Kawahara, A.; Shimamoto, N.; Muneoka, Y.
Brain Res. 422, 374-376, 1987
A;Title: Catch-relaxing peptide isolated from Mytilus pedal ganglia.
A;Reference number: A29342; MUID:88052022
A;Accession: A29342
A;Molecule type: protein
A;Residues: 1-7 <HIR>
C;Comment: This peptide exhibits both potentiating (contraction) and inhibitory (relaxat
C;Superfamily: unassigned animal peptides
C;Keywords: amidated carboxyl end; hormone; retractor muscle
F;7/Modified site: amidated carboxyl end (Leu) #status experimental

  Query Match             35.1%;  Score 13;  DB 2;  Length 7;
  Best Local Similarity   66.7%;  Pred. No. 2.2e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      4 VPM 6
          :||
Db      2 MPM 4


RESULT   5
A28340
myomodulin - California sea hare
C;Species: Aplysia californica (California sea hare)
C;Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 18-Jun-1993
C;Accession: A28340
R;Cropper, E.C.; Tenenbaum, R.; Kolks, M.A.G.; Kupfermann, I.; Weiss, K.R.
Proc. Natl. Acad. Sci. U.S.A. 84, 5483-5486, 1987
A;Title: Myomodulin: A bioactive neuropeptide present in an identified cholinergic bucca
A;Reference number: A28340; MUID:87261010
A;Accession: A28340
A;Molecule type: protein
A;Residues: 1-7 <CRO>

  Query Match             35.1%;  Score 13;  DB 2;  Length 7;
  Best Local Similarity   66.7%;  Pred. No. 2.2e+05;
  Matches    2;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      5 PMN 7
          ||:
Db      1 PMS 3


RESULT   6
A27696
contraction-inhibiting peptide I - blue mussel
C;Species: Mytilus edulis (blue mussel)
C;Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 08-Dec-1995
C;Accession: A27696
R;Hirata, T.; Kubota, I.; Iwasawa, N.; Takabatake, I.; Ikeda, T.; Muneoka, Y.
Biochem. Biophys. Res. Commun. 152, 1376-1382, 1988
A;Title: Structures and actions of Mytilus inhibitory peptides.
A;Reference number: A90142; MUID:88240357
A;Accession: A27696
A;Molecule type: protein
A;Residues: 1-6 <HIR>
C;Keywords: amidated carboxyl end
F;6/Modified site: amidated carboxyl end (Val) #status experimental

  Query Match             32.4%;  Score 12;  DB 2;  Length 6;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      5 PM 6
          ||
Db      3 PM 4


RESULT   7
B27696
contraction-inhibiting peptide II - blue mussel
C;Species: Mytilus edulis (blue mussel)
C;Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 08-Dec-1995
C;Accession: B27696
R;Hirata, T.; Kubota, I.; Iwasawa, N.; Takabatake, I.; Ikeda, T.; Muneoka, Y.
Biochem. Biophys. Res. Commun. 152, 1376-1382, 1988
A;Title: Structures and actions of Mytilus inhibitory peptides.
A;Reference number: A90142; MUID:88240357
A;Accession: B27696
A;Molecule type: protein
A;Residues: 1-6 <HIR>
C;Keywords: amidated carboxyl end
F;6/Modified site: amidated carboxyl end (Val) #status experimental

  Query Match             32.4%;  Score 12;  DB 2;  Length 6;
  Best Local Similarity   100.0%;  Pred. No. 2.2e+05;
  Matches    2;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      5 PM 6
          ||
Db      3 PM 4


RESULT   8
A61300
22K superhelical DNA-binding protein - Escherichia coli (fragment)
C;Species: Escherichia coli
C;Date: 17-Jul-1994 #sequence_revision 17-Jul-1994 #text_change 07-May-1999
C;Accession: A61300
R;Kishi, F.; Ebina, Y.; Miki, T.; Nakazawa, T.; Nakazawa, A.
J. Biochem. 92, 1059-1068, 1982
A;Title: Purification and characterization of a protein from Escherichia coli which f
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: January 6, 2002, 11:07:52 ; Search time 79.01 Seconds
(without alignments)
68.795 Million cell updates/sec

Title: US-09-897-042-13
Perfect score: 24
Sequence: 1 atgaacttggtaccgatgaactaa 24

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 281400

Minimum DB seq length: 0
Maximum DB seq length: 24

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	13.6	56.7	20	4	US-09-177-249-316	Sequence 316, App
	2	12.4	51.7	18	3	US-08-438-506-5	Sequence 5, Appli
	3	12.4	51.7	20	1	US-08-325-071-52	Sequence 52, Appl
	4	12.4	51.7	20	4	US-08-461-004A-52	Sequence 52, Appl
	5	12.4	51.7	21	1	US-08-325-071-54	Sequence 54, Appl
	6	12.4	51.7	21	4	US-08-461-004A-54	Sequence 54, Appl
c	7	12.4	51.7	24	1	US-07-783-861C-13	Sequence 13, Appl
c	8	12	50.0	24	2	US-08-805-918-19	Sequence 19, Appl
c	9	12	50.0	24	2	US-08-805-918-22	Sequence 22, Appl
c	10	12	50.0	24	2	US-08-811-028-15	Sequence 15, Appl
c	11	11.6	48.3	19	1	US-08-110-294A-44	Sequence 44, Appl
c	12	11.6	48.3	19	2	US-08-389-926-44	Sequence 44, Appl
	13	11.6	48.3	20	2	US-08-117-952-82	Sequence 82, Appl
	14	11.6	48.3	21	1	US-08-075-533-15	Sequence 15, Appl
	15	11.6	48.3	21	1	US-08-125-012-18	Sequence 18, Appl
	16	11.6	48.3	21	1	US-08-221-579A-12	Sequence 12, Appl
	17	11.6	48.3	21	1	US-08-783-818-18	Sequence 18, Appl
	18	11.6	48.3	21	2	US-08-453-349-18	Sequence 18, Appl
	19	11.6	48.3	21	2	US-08-948-176-15	Sequence 15, Appl
	20	11.6	48.3	21	2	US-08-704-701-12	Sequence 12, Appl
	21	11.6	48.3	21	2	US-08-979-385B-16	Sequence 16, Appl
	22	11.6	48.3	21	4	US-09-321-461-12	Sequence 12, Appl
	23	11.6	48.3	21	5	PCT-US91-09160-15	Sequence 15, Appl
	24	11.6	48.3	21	5	PCT-US95-16206A-4	Sequence 4, Appli
c	25	11.6	48.3	24	1	US-08-316-293-65	Sequence 65, Appl
	26	11.6	48.3	24	1	US-08-316-293-66	Sequence 66, Appl
c	27	11.4	47.5	16	1	US-08-325-071-51	Sequence 51, Appl
c	28	11.4	47.5	16	4	US-08-461-004A-51	Sequence 51, Appl
c	29	11.4	47.5	18	3	US-08-849-075-6	Sequence 6, Appli
c	30	11.4	47.5	19	3	US-09-175-219A-1	Sequence 1, Appli
	31	11.4	47.5	20	1	US-08-096-182A-13	Sequence 13, Appl
	32	11.4	47.5	20	1	US-08-877-109-13	Sequence 13, Appl
	33	11.4	47.5	20	2	US-08-734-941-3	Sequence 3, Appli
	34	11.4	47.5	20	3	US-08-798-760-13	Sequence 13, Appl
	35	11.4	47.5	20	5	PCT-US94-08327-13	Sequence 13, Appl
	36	11.4	47.5	21	1	US-08-586-274-4	Sequence 4, Appli
	37	11.4	47.5	21	1	US-08-586-274-24	Sequence 24, Appl
	38	11.4	47.5	21	3	US-09-116-294-15	Sequence 15, Appl
	39	11.4	47.5	21	3	US-08-849-075-2	Sequence 2, Appli
	40	11.4	47.5	21	4	US-09-537-357-50	Sequence 50, Appl
c	41	11.4	47.5	22	4	US-08-949-155-15	Sequence 15, Appl
	42	11.4	47.5	24	1	US-07-783-861C-12	Sequence 12, Appl
	43	11.4	47.5	24	6	5256545-41	Patent No. 5256545
	44	11.2	46.7	17	4	US-09-068-860-4	Sequence 4, Appli
	45	11.2	46.7	17	4	US-09-068-860-12	Sequence 12, Appl
c	46	11.2	46.7	21	1	US-08-075-533-16	Sequence 16, Appl
c	47	11.2	46.7	21	2	US-08-948-176-16	Sequence 16, Appl
c	48	11.2	46.7	21	4	US-09-050-159-60	Sequence 60, Appl
c	49	11.2	46.7	21	5	PCT-US91-09160-16	Sequence 16, Appl
	50	11.2	46.7	22	1	US-08-717-526-21	Sequence 21, Appl

ALIGNMENTS

RESULT 1
US-09-177-249-316/c
; Sequence 316, Application US/09177249
; Patent No. 6229064
; GENERAL INFORMATION:
; APPLICANT: Fischer, Robert L.
; APPLICANT: Ohad, Nir
; APPLICANT: Kiyosue, Tomohiro
; APPLICANT: Yadegari, Ramin
; APPLICANT: Margossian, Linda
; APPLICANT: Harada, John
; APPLICANT: Goldberg, Robert B.
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Nucleic Acids That Control Seed and Fruit
; TITLE OF INVENTION: Development in Plants
; FILE REFERENCE: 023070-086120US
; CURRENT APPLICATION NUMBER: US/09/177,249
; CURRENT FILING DATE: 1998-10-22
; EARLIER APPLICATION NUMBER: US 09/071,838
; EARLIER FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 324
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 316
; LENGTH: 20
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:primer
; OTHER INFORMATION: cerln1570
US-09-177-249-316

Query Match 56.7%; Score 13.6; DB 4; Length 20;
Best Local Similarity 80.0%; Pred. No. 5.1e+02;
Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
Qy       3 gaacttggtaccgatgaact 22
           ||||||  || ||||||  ||
Db      20 GAACTTCATAGCGATGACCT 1
```

RESULT 2
US-08-438-506-5
; Sequence 5, Application US/08438506
; Patent No. 6001562

```
;  GENERAL INFORMATION:
;    APPLICANT:  Milosavljevic, Aleksandar
;    TITLE OF INVENTION:  DNA SEQUENCE SIMILARITY RECOGNITION BY
;    TITLE OF INVENTION:  HYBRIDIZATION TO SHORT OLIGOMERS
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  United States of America
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS/ASCII
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/438,506
;      FILING DATE:  Concurrently Herewith
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Highlander, Steven L.
;      REGISTRATION NUMBER:  37,642
;      REFERENCE/DOCKET NUMBER:  ARCD:153/HYL
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (512) 418-3000
;      TELEFAX:  (713) 789-2679
;      TELEX:  79-0924
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  18 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-438-506-5


  Query Match             51.7%;  Score 12.4;  DB 3;  Length 18;
  Best Local Similarity   92.9%;  Pred. No. 1.9e+03;
  Matches   13;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      11 taccgatgaactaa 24
           ||||||||| ||||
Db       4 TACCGATGAGCTAA 17


ESULT   3
US-08-325-071-52
; Sequence 52, Application US/08325071
; Patent No. 5587311
;  GENERAL INFORMATION:
;    APPLICANT:  COBON, Stewart Gary
;    APPLICANT:  MOORE, Joanna Terry
;    APPLICANT:  JOHNSON, Law Anthony York
;    APPLICANT:  WILLADSEN, Peter
;    APPLICANT:  KEMP, David Harold
;    APPLICANT:  SRISKANTHA, Alagacone
;    APPLICANT:  RIDING, George Alfred
;    APPLICANT:  RAND, Keith No. 5587311man
;    TITLE OF INVENTION:  DNA Encoding A Cell Membrane
;    TITLE OF INVENTION:  Glycoprotein Of A Tick Gut
;    NUMBER OF SEQUENCES:  71
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W.
;      CITY:  Washington, D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/325,071
;      FILING DATE:  14-OCT-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/062,109
;      FILING DATE:  17-MAY-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/926,368
;      FILING DATE:  07-AUG-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/242,196
;      FILING DATE:  06-JUL-1988
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/AU87/00401
;      FILING DATE:  27-NOV-1987
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  AU PI4912
;      FILING DATE:  16-OCT-1987
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  AU PI2570
;      FILING DATE:  19-JUN-1987
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  AU PH9196
;      FILING DATE:  27-NOV-1986
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  60042/111 BIAU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202 672 5300
;      TELEFAX:  202 672 5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  52:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  20 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    ANTI-SENSE:  YES
US-08-325-071-52


  Query Match             51.7%;  Score 12.4;  DB 1;  Length 20;
  Best Local Similarity   92.9%;  Pred. No. 1.9e+03;
  Matches   13;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 gaacttggtaccga 16
           ||||| ||||||||
Db       4 GAACTGGGTACCGA 17


RESULT   4
US-08-461-004A-52
; Sequence 52, Application US/08461004A
; Patent No. 6235283
;  GENERAL INFORMATION:
;    APPLICANT:  COBON, Stewart Gary
;    APPLICANT:  MOORE, Joanna Terry
;    APPLICANT:  JOHNSON, Law Anthony York
;    APPLICANT:  WILLADSEN, Peter
;    APPLICANT:  KEMP, David Harold
;    APPLICANT:  SRISKANTHA, Alagacone
;    APPLICANT:  RIDING, George Alfred
;    APPLICANT:  RAND, Keith No. 6235283man
;    TITLE OF INVENTION:  DNA Encoding A Cell Membrane
;    TITLE OF INVENTION:  Glycoprotein Of A Tick Gut
;    NUMBER OF SEQUENCES:  71
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W.
;      CITY:  Washington, D.C.
;      COUNTRY:  USA
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: January 6, 2002, 12:24:52 ; Search time 142.39 Seconds
(without alignments)
144.503 Million cell updates/sec

Title: US-09-897-042-13
Perfect score: 24
Sequence: 1 atgaacttggtaccgatgaactaa 24

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 607738

Minimum DB seq length: 0
Maximum DB seq length: 24

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : N_Geneseq_1101:*
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	13	54.2	20	18	AAT60152	Primer #5 for tyro
c	2	12.8	53.3	24	19	AAV34719	Human Fanconi I pr
c	3	12.6	52.5	20	21	AAA58055	Human interleukin-
c	4	12.6	52.5	22	20	AAZ40501	Human STE20-relate
c	5	12.6	52.5	24	20	AAX08927	Primer for modifie
	6	12.4	51.7	18	21	AAZ46447	Target sequence ID
c	7	12.2	50.8	17	20	AAA17376	Aryl hydrocarbon n
c	8	12.2	50.8	18	22	AAC99290	Probe sequence use
c	9	12.2	50.8	20	18	AAT89747	Alpha amylase A ge
	10	12.2	50.8	20	20	AAX96327	PCR primer used to
	11	12	50.0	20	22	AAF75655	Murine ztrypl codi
	12	12	50.0	21	12	AAQ15384	Toxigenic Clostrid
	13	12	50.0	21	19	AAV64140	Rat activin recept
c	14	11.8	49.2	17	20	AAX32175	Probe specific for
	15	11.8	49.2	19	21	AAA83059	cdk6 ribozyme bind
	16	11.8	49.2	19	22	AAH58221	Cell-cycle depende
c	17	11.8	49.2	20	20	AAX93833	PCR primer used to
	18	11.8	49.2	21	22	AAH62391	Voltage gated Ca c
c	19	11.8	49.2	24	22	AAF63472	Oligonucleotide CL
c	20	11.6	48.3	19	18	AAT77582	Wheat microsatelli
	21	11.6	48.3	20	16	AAQ82082	Chromosome 11 (loc
	22	11.6	48.3	21	13	AAQ26534	Acyl-ACP thioester
	23	11.6	48.3	21	15	AAQ67153	Primer for amplify
	24	11.6	48.3	21	16	AAT00841	HIV gag/pol gene j
	25	11.6	48.3	21	17	AAT32899	HIV-1 HXB2 gag/pol
	26	11.6	48.3	21	18	AAT77580	Wheat microsatelli
	27	11.6	48.3	21	20	AAX80893	Acyl-ACP thioester
c	28	11.6	48.3	21	21	AAA94143	Adenovirus 5 E2/E3
	29	11.6	48.3	21	22	AAF96404	Human gene single
c	30	11.6	48.3	22	21	AAA96903	RACE PCR primer us
	31	11.6	48.3	22	22	AAS03611	PCR primer rend9b1
c	32	11.6	48.3	23	20	AAX26142	Primer for generat
c	33	11.6	48.3	24	13	AAQ23101	HPV11 probe LCR2B.
	34	11.6	48.3	24	13	AAQ23102	HPV11 probe LCR2B'
c	35	11.6	48.3	24	17	AAT29890	Human papillomavir
	36	11.6	48.3	24	17	AAT29891	Human papillomavir
c	37	11.4	47.5	17	20	AAA17377	Aryl hydrocarbon n
c	38	11.4	47.5	19	21	AAZ73913	Human biallelic ma
	39	11.4	47.5	20	16	AAQ85297	5' oligo primer fo
	40	11.4	47.5	20	18	AAT60151	Primer #4 for tyro
	41	11.4	47.5	20	18	AAT60154	Primer #7 for tyro
	42	11.4	47.5	20	18	AAT60153	Primer #6 for tyro
	43	11.4	47.5	20	20	AAX01097	IgA-binding fragme
	44	11.4	47.5	21	17	AAT49325	Methicillin resist
c	45	11.4	47.5	22	19	AAV29168	Nucleotide sequenc
c	46	11.4	47.5	24	21	AAZ87485	Human ADH7 gene ex
c	47	11.4	47.5	24	22	AAH46732	Type 11 phosphodie
	48	11.2	46.7	17	20	AAV71940	Primer used for se
	49	11.2	46.7	17	21	AAA58620	Primer used to pro
c	50	11.2	46.7	18	13	AAQ22881	HCV-Hc59 primer #1

ALIGNMENTS

RESULT 1
AAT60152
ID AAT60152 standard; DNA; 20 BP.
XX
AC AAT60152;
XX
DT 25-NOV-1997 (first entry)
XX
DE Primer #5 for tyrosine 3-hydroxylase gene from vitelline cell cDNA.
XX
KW Primer; amplify; schistosome; tyrosine 3-hydroxylase; TY3H; vitelline;
KW drosophila; transgenic schistosome; integument; Mehlis glands; PCR;
KW ootype; pronuclei; miracidia; snail; cercarium; insulin; leptin;
KW calcitonin; alpha 1 anti-trypsin; factor VIII; HIV co-receptor ligand;
KW cholesterol ester transfer protein inhibitor; gene therapy; ss.
XX
OS Synthetic.
XX
PN WO9711191-A1.
XX
PD 27-MAR-1997.
XX
PF 20-SEP-1996; 96WO-US15083.
XX
PR 21-SEP-1995; 95US-0004115.
XX
PA (MILL/) MILLER I.
XX
PI Miller I;

```
XX
DR   WPI; 1997-202898/18.
XX
PT   Secretion of therapeutic proteins - from transgenic schistosomes
PT   into the blood of a animal host
XX
PS   Disclosure; Page 37; 48pp; English.
XX
CC   The sequences given in AAT60148-54 are primers which were used in the
CC   amplification of the schistosome tyrosine 3-hydroxylase (TY3H) from
CC   vitelline cell cDNA. The upstream primers (AAT60148-50) are based on
CC   drosophila TY3H amino acids 368-373, and the downstream primers
CC   (AAT60151-54) are based on drosophila TY3H amino acids 428-433. The
CC   amplified sequences may be used in the method of the invention for the
CC   production of transgenic schistosomes for secretion of particular gene
CC   products. The method comprises cloning the protein coding portion of a
CC   cDNA corresponding to the gene encoding the protein of interest into a
CC   plasmid having upstream and downstream promoter/enhancer sequences, and
CC   containing 5' and 3' untranslated regions of a schistosome gene that is
CC   expressed in the integument, vitelline or Mehlis glands, or ootype of
CC   schistosomes. The transgene DNA is microinjected into the pronucleii or
CC   cytoplasm of the zygotes of stage I schistosome eggs and the eggs are
CC   cultured to maturity in vitro. Miracidia are hatched and snails are
CC   infected with one miracidium each. The snails are cultured until
CC   cercaria of schistosome clones are produced. Transgenic clones are
CC   identified at the cercarial stage and maintained by sporocyst transfer.
CC   Laboratory animals are infected with one cercarium each. Transgenic
CC   schistosome clones that secrete high levels of the desired protein
CC   into the peripheral blood of the host are then identified. The method
CC   is used to produce therapeutic proteins in the blood stream of animals.
CC   Typical proteins to be produced include insulin, leptin, calcitonin,
CC   alpha 1 anti-trypsin, factor VIII, HIV co-receptor ligands and
CC   cholesterol ester transfer protein inhibitor. Use of schistosomes as
CC   intermediate vectors for gene therapy facilitates mass production,
CC   quality control, termination of treatment and dose titration. It avoids
CC   problems, such as low efficiency, possible infection by contaminating
CC   replication competent virus, potential recombination with host DNA,
CC   possible malignant transformation and need to individualise treatments,
CC   associated with use of viral vectors.
XX
SQ   Sequence 20 BP; 3 A; 2 C; 5 G; 5 T; 5 other;


  Query Match             54.2%;  Score 13;  DB 18;  Length 20;
  Best Local Similarity   65.0%;  Pred. No. 2.8e+03;
  Matches   13;  Conservative    3;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 tgaacttggtaccgatgaac 21
           ||: |:||:||  | |||||
Db       1 tgrtcytgrtanggytgaac 20


RESULT   2
AAV34719/c
ID   AAV34719 standard; cDNA; 24 BP.
XX
AC   AAV34719;
XX
DT   27-AUG-1998  (first entry)
XX
DE   Human Fanconi I primer SP2.
XX
KW   Fanconi; anaemia; disease; diagnosis; disorder; predisposition; tumour;
KW   cell cycle; cell activation; DNA repair; cytopaenia; treatment; effector;
KW   prevention; gene therapy; proliferative disease; therapeutic agent;
KW   primer; ss.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9816637-A1.
XX
PD   23-APR-1998.
XX
PF   08-OCT-1997;   97WO-EP05543.
XX
PR   11-OCT-1996;   96EP-0116336.
XX
PA   (BOEF ) BOEHRINGER MANNHEIM GMBH.
XX
PI   Kubbies M,  Machl A,  Planitzer S;
XX
DR   WPI; 1998-261040/23.
XX
PT   Nucleic acid corresponding to Fanconi anaemia gene - used for
PT   diagnosis, treatment and prevention of diseases involving disordered
PT   cell cycle progression, cytopaenia etc.
XX
PS   Example 4; Page 15; 37pp; German.
XX
CC   AAV34718-V34722 are primers used in the isolation of a gene associated
CC   with Fanconi anaemia. This sequence can be used for diagnosing disease,
CC   or predisposition to it, associated with disorders of cell cycling,
CC   activation or cell-cycle progression; DNA repair; cytopaenia; or
CC   formation/progression of tumours, or to treat or prevent such
CC   conditions, particularly by gene therapy, especially of cytopaenia,
CC   tumours or other proliferative diseases. Cells that express this protein
CC   can be used to identify specific effectors (potentially useful as
CC   therapeutic agents).
XX
SQ   Sequence 24 BP; 8 A; 6 C; 5 G; 5 T; 0 other;


  Query Match             53.3%;  Score 12.8;  DB 19;  Length 24;
  Best Local Similarity   87.5%;  Pred. No. 3.6e+03;
  Matches   14;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       7 ttggtaccgatgaact 22
           ||||||| ||||| ||
Db      24 TTGGTACAGATGATCT 9


RESULT   3
AAA58055/c
ID   AAA58055 standard; DNA; 20 BP.
XX
AC   AAA58055;
XX
DT   10-OCT-2000  (first entry)
XX
DE   Human interleukin-15 (IL-15) antisense oligonucleotide, ODN 10.
XX
KW   Human interleukin-15; IL-15; antisense oligonucleotide;
KW   expression inhibition; T-cell mediated immune response;
KW   autoimmune disorder; inflammatory polyarthropathy; rheumatoid arthritis;
KW   transplant rejection; graft versus host disease; lupus erythematosus;
KW   asthma; inflammatory bowel disease; gene therapy; immunosuppressive;
KW   antiinflammatory; 3' UTR; 3' untranslated region; ss.
XX
OS   Homo sapiens.
XX
PN   WO200028019-A2.
XX
PD   18-MAY-2000.
XX
PF   04-NOV-1999;   99WO-IL00589.
XX
PR   05-NOV-1998;   98IL-0126919.
XX
PA   (UYNE ) UNIV BEN-GURION NEGEV RES & DEV.
PA   (MORR-) MOR RES APPL LTD.
XX
PI   Douvdevani A,  Chaimovitz C;
XX
DR   WPI; 2000-376536/32.
XX
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: January 6, 2002, 11:01:22 ; Search time 1365.98 Seconds
(without alignments)
188.801 Million cell updates/sec

Title: US-09-897-042-13
Perfect score: 24
Sequence: 1 atgaacttggtaccgatgaactaa 24

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 13210

Minimum DB seq length: 0
Maximum DB seq length: 24

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_htc:*
10: gb_est1:*
11: gb_est2:*
12: gb_htc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	11.2	46.7	24	13	AZ581676	AZ581676 1M0370N06
	2	10.8	45.0	19	13	AZ340577	AZ340577 1M0072H21
c	3	10.8	45.0	21	13	AZ663981	AZ663981 1M0543J22
	4	10.2	42.5	22	13	AZ501093	AZ501093 1M0339L13
c	5	10.2	42.5	24	13	AZ360525	AZ360525 1M0103N13
	6	10	41.7	19	10	AU061154	AU061154 AU061154
	7	10	41.7	20	13	TA339H11Q	AL492631 T. brucei
	8	9.8	40.8	20	13	AZ620235	AZ620235 1M0452E23
c	9	9.8	40.8	22	13	AZ779213	AZ779213 2M0015L18
	10	9.6	40.0	21	13	AZ770212	AZ770212 1M0571G10
	11	9.6	40.0	24	13	TA116E02P	AL462525 T. brucei
	12	9.4	39.2	19	13	AZ836789	AZ836789 2M0131I18
c	13	9.4	39.2	21	13	AZ828967	AZ828967 2M0106N15
	14	9.4	39.2	21	13	TA20G06P	AL454552 T. brucei
c	15	9.4	39.2	22	13	TA114G12Q	AL462979 T. brucei
c	16	9.4	39.2	23	13	AZ333221	AZ333221 1M0062O11
	17	9.4	39.2	23	13	AZ459779	AZ459779 1M0264A22
	18	9.2	38.3	17	2	HSM007757	A1042907 Homo sapi
	19	9.2	38.3	19	13	AZ787717	AZ787717 2M0034A21
	20	9.2	38.3	21	13	AZ609424	AZ609424 1M0434O16
	21	9.2	38.3	21	13	AZ828967	AZ828967 2M0106N15
	22	9.2	38.3	22	13	AZ394329	AZ394329 1M0157B21
	23	9	37.5	19	13	AZ399413	AZ399413 1M0165E16
c	24	9	37.5	19	13	AZ815827	AZ815827 2M0084K23
	25	9	37.5	19	13	AZ991317	AZ991317 2M0275K24
c	26	9	37.5	20	13	AZ658617	AZ658617 1M0535E04
	27	9	37.5	21	13	AZ768984	AZ768984 1M0569C16
	28	9	37.5	22	13	AZ635609	AZ635609 1M0493C08
c	29	9	37.5	22	13	AZ840373	AZ840373 2M0136G21
	30	9	37.5	22	13	AZ871339	AZ871339 2M0184G09
	31	9	37.5	23	13	AZ308507	AZ308507 1M0011L02
	32	9	37.5	24	13	TA294D03Q	AL486340 T. brucei
	33	8.8	36.7	19	13	AZ831033	AZ831033 2M0110O01
c	34	8.8	36.7	22	13	AZ429443	AZ429443 1M0213G21
c	35	8.8	36.7	22	13	AZ450734	AZ450734 1M0249B10
	36	8.8	36.7	22	13	AZ589849	AZ589849 1M0399A09
	37	8.8	36.7	22	13	TA19D05Q	AL452318 T. brucei
	38	8.8	36.7	23	13	AZ370804	AZ370804 1M0121B20
	39	8.8	36.7	23	13	AZ828383	AZ828383 2M0105O15
c	40	8.8	36.7	24	13	AZ313141	AZ313141 1M0029D19
	41	8.8	36.7	24	13	AZ342545	AZ342545 1M0075I11
	42	8.8	36.7	24	13	AZ804947	AZ804947 2M0066B07
	43	8.6	35.8	20	13	AZ936985	AZ936985 2M0195E06
c	44	8.6	35.8	20	13	AZ943793	AZ943793 2M0204A07
	45	8.6	35.8	21	13	AZ453408	AZ453408 1M0254H12
c	46	8.6	35.8	22	13	AZ394329	AZ394329 1M0157B21
	47	8.6	35.8	23	13	AZ805705	AZ805705 2M0067D18
	48	8.6	35.8	23	13	TA195D06Q	AL476956 T. brucei
c	49	8.6	35.8	24	13	AZ628737	AZ628737 1M0481E09
c	50	8.4	35.0	19	13	AZ312945	AZ312945 1M0029P03

ALIGNMENTS

RESULT 1
AZ581676/c
LOCUS AZ581676 24 bp DNA GSS 13-DEC-2000
DEFINITION 1M0370N06R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0370N06 R, DNA sequence.
ACCESSION AZ581676
VERSION AZ581676.1 GI:11696926
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 24)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0370 row: N column: 06

```
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 24.
FEATURES              Location/Qualifiers
     source           1. .24
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="UUGC1M0370N06"
                      /clone_lib="Mouse 10kb plasmid UUGC1M library"
                      /sex="Male"
                      /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                      /note="Vector: PWD42nv; Purified genomic DNA from M.
                      musculus C57BL/6J (male) was obtained  from the Jackson
                      Laboratory Mouse DNA Resource
                      (http://www.jax.org/resources/documents/dnares/). The DNA
                      was hydrodynamically sheared by repeated passage through a
                      0.005 inch orifice at constant velocity. The sheared DNA
                      was blunt end-repaired with T4 DNA polymerase and T4
                      polynucleotide kinase. Adaptor oligonucleotides were
                      ligated to the blunt ends in high  molar excess. The
                      adaptored DNA was purified and size-selected for a 9.5 to
                      10.5 kb range using preparative agarose gel
                      electrophoresis. Vector DNA was prepared from a derivative
                      of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                      inducible derivative of plasmid R1. The vector was ligated
                      with adaptors complementary to the insert adaptors and
                      purified. The sheared, adaptored mouse DNA was annealed to
                      adaptored vector DNA, and transformed into
                      chemically-competent E. coli XL10-Gold (Stratagene) cells
                      and selected for ampicillin resistance."
BASE COUNT        5 a      4 c      4 g      11 t
ORIGIN


  Query Match           46.7%;  Score 11.2;  DB 13;  Length 24;
  Best Local Similarity  81.2%;  Pred. No. 3.8e+05;
  Matches   13;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       9 ggtaccgatgaactaa 24
           | || ||||||| |||
Db      18 GCTATCGATGAATTAA 3


RESULT   2
AZ340577
LOCUS       AZ340577       19 bp    DNA             GSS       29-SEP-2000
DEFINITION  1M0072H21R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0072H21 R, DNA sequence.
ACCESSION   AZ340577
VERSION     AZ340577.1  GI:10415969
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 19)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0072  row: H  column: 21
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 19.
FEATURES              Location/Qualifiers
     source           1. .19
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="UUGC1M0072H21"
                      /clone_lib="Mouse 10kb plasmid UUGC1M library"
                      /sex="Male"
                      /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                      /note="Vector: PWD42nv; Purified genomic DNA from M.
                      musculus C57BL/6J (male) was obtained  from the Jackson
                      Laboratory Mouse DNA Resource
                      (http://www.jax.org/resources/documents/dnares/). The DNA
                      was hydrodynamically sheared by repeated passage through a
                      0.005 inch orifice at constant velocity. The sheared DNA
                      was blunt end-repaired with T4 DNA polymerase and T4
                      polynucleotide kinase. Adaptor oligonucleotides were
                      ligated to the blunt ends in high  molar excess. The
                      adaptored DNA was purified and size-selected for a 9.5 to
                      10.5 kb range using preparative agarose gel
                      electrophoresis. Vector DNA was prepared from a derivative
                      of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                      inducible derivative of plasmid R1. The vector was ligated
                      with adaptors complementary to the insert adaptors and
                      purified. The sheared, adaptored mouse DNA was annealed to
                      adaptored vector DNA, and transformed into
                      chemically-competent E. coli XL10-Gold (Stratagene) cells
                      and selected for ampicillin resistance."
BASE COUNT        5 a      4 c      6 g      4 t
ORIGIN


  Query Match           45.0%;  Score 10.8;  DB 13;  Length 19;
  Best Local Similarity  85.7%;  Pred. No. 5.7e+05;
  Matches   12;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       9 ggtaccgatgaact 22
           |||| || ||||||
Db       3 GGTAACGCTGAACT 16


RESULT   3
AZ663981/c
LOCUS       AZ663981       21 bp    DNA             GSS       14-DEC-2000
DEFINITION  1M0543J22R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0543J22 R, DNA sequence.
ACCESSION   AZ663981
VERSION     AZ663981.1  GI:11801127
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 21)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: January 6, 2002, 11:02:32 ; Search time 1434.12 Seconds
(without alignments)
276.080 Million cell updates/sec

Title: US-09-897-042-13
Perfect score: 24
Sequence: 1 atgaacttggtaccgatgaactaa 24

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 309068

Minimum DB seq length: 0
Maximum DB seq length: 24

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vi:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rod:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rod:*
36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	Query Match	Length	DB	ID	Description
c	1	13.6	56.7	20	6	AR150940	AR150940 Sequence
c	2	12.6	52.5	20	6	AX024726	AX024726 Sequence
	3	12.4	51.7	18	6	AR093435	AR093435 Sequence
	4	12.4	51.7	20	6	AR152501	AR152501 Sequence
	5	12.4	51.7	20	6	I32378	I32378 Sequence 52
	6	12.4	51.7	21	6	AR152503	AR152503 Sequence
	7	12.4	51.7	21	6	I32380	I32380 Sequence 54
c	8	12.4	51.7	24	6	I15196	I15196 Sequence 13
	9	12	50.0	20	6	AX082987	AX082987 Sequence
	10	12	50.0	21	6	AX029243	AX029243 Sequence
c	11	12	50.0	24	6	AR069541	AR069541 Sequence
c	12	12	50.0	24	6	E16933	E16933 Linker DNA.
c	13	12	50.0	24	6	E16936	E16936 Linker DNA.
	14	11.8	49.2	19	6	AX129427	AX129427 Sequence
c	15	11.8	49.2	21	6	AX095881	AX095881 Sequence
c	16	11.8	49.2	24	6	AX085594	AX085594 Sequence
c	17	11.6	48.3	18	6	A35582	A35582 Synthetic h
c	18	11.6	48.3	19	6	AR034541	AR034541 Sequence
c	19	11.6	48.3	19	6	AR048641	AR048641 Sequence
	20	11.6	48.3	20	6	AR066734	AR066734 Sequence
	21	11.6	48.3	21	6	AR001369	AR001369 Sequence
	22	11.6	48.3	21	6	AR044856	AR044856 Sequence
	23	11.6	48.3	22	6	AR052242	AR052242 Sequence
	24				6	AR078349	AR078349 Sequence
					6	AR085200	AR085200 Sequence
					6	AR138120	AR138120 Sequence
					6	AX036805	AX036805 Sequence
					6	I33906	I33906 Sequence 18
					6	I17054	I17054 Sequence 65
					6	I17055	I17055 Sequence 66
					6	AR152500	AR152500 Sequence
					6	I32377	I32377 Sequence 51
					6	AR110519	AR110519 Sequence
					6	AR004782	AR004782 Sequence
					6	AR032321	AR032321 Sequence
					6	I13582	I13582 Sequence 13
					6	AX096502	AX096502 Sequence
					6	AX154194	AX154194 Sequence
			47.5	21	6	E12099	E12099 PCR primer
	40	11.4	47.5	21	6	I86678	I86678 Sequence 4
	41	11.4	47.5	21	6	I86698	I86698 Sequence 24
	42	11.4	47.5	23	6	AX067767	AX067767 Sequence
	43	11.4	47.5	24	6	I15195	I15195 Sequence 12
c	44	11.2	46.7	20	6	E15729	E15729 PCR primer.
c	45	11.2	46.7	21	6	AR137313	AR137313 Sequence
c	46	11.2	46.7	21	6	AX037435	AX037435 Sequence
	47	11.2	46.7	21	6	AX095991	AX095991 Sequence
	48	11.2	46.7	22	6	A61184	A61184 Sequence 21
	49	11.2	46.7	24	6	A76352	A76352 Sequence 18
c	50	11	45.8	18	6	AR031334	AR031334 Sequence

closed

15 not run — too large

ALIGNMENTS

RESULT 1
AR150940/c
LOCUS AR150940 20 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 316 from patent US 6229064.
ACCESSION AR150940
VERSION AR150940.1 GI:15115531
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 20)
AUTHORS Fischer,R.L., Ohad,N., Kiyosue,T., Yadegari,R., Margossian,L., Harada,J. and Goldberg,R.B.
TITLE Nucleic acids that control endosperm development in plants
JOURNAL Patent: US 6229064-A 316 08-MAY-2001;

```
FEATURES             Location/Qualifiers
     source          1. .20
                     /organism="unknown"
BASE COUNT        5 a      4 c      5 g      6 t
ORIGIN


  Query Match             56.7%;  Score 13.6;  DB 6;  Length 20;
  Best Local Similarity   80.0%;  Pred. No. 4e+04;
  Matches   16;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      3 gaacttggtaccgatgaact 22
          ||||||  || ||||||  ||
Db     20 GAACTTCATAGCGATGACCT 1


RESULT   2
AX024726/c
LOCUS       AX024726       20 bp    DNA             PAT       15-SEP-2000
DEFINITION  Sequence 12 from Patent WO0028019.
ACCESSION   AX024726
VERSION     AX024726.1  GI:10184805
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 20)
  AUTHORS   Douvdevani,A. and Chaimovitz,C.
  TITLE     Antisense oligomer
  JOURNAL   Patent: WO 0028019-A 12 18-MAY-2000;
            MOR RESEARCH APPLIC LTD (IL) ; DOUVDEVANI AMOS (IL) ; UNIV BEN
            GURION (IL) ; CHAIMOVITZ CIDIO (IL)
FEATURES             Location/Qualifiers
     source          1. .20
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT        5 a      4 c      5 g      6 t
ORIGIN


  Query Match             52.5%;  Score 12.6;  DB 6;  Length 20;
  Best Local Similarity   78.9%;  Pred. No. 1.3e+05;
  Matches   15;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      4 aacttggtaccgatgaact 22
          | |||||  | ||||||||
Db     20 ACCTTGGATCAGATGAACT 2


RESULT   3
AR093435
LOCUS       AR093435       18 bp    DNA             PAT       08-SEP-2000
DEFINITION  Sequence 5 from patent US 6001562.
ACCESSION   AR093435
VERSION     AR093435.1  GI:10020184
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 18)
  AUTHORS   Milosavljevic,A.
  TITLE     DNA sequence similarity recognition by hybridization to short
            oligomers
  JOURNAL   Patent: US 6001562-A 5 14-DEC-1999;
FEATURES             Location/Qualifiers
     source          1. .18
                     /organism="unknown"
BASE COUNT        6 a      3 c      4 g      5 t
ORIGIN


  Query Match             51.7%;  Score 12.4;  DB 6;  Length 18;
  Best Local Similarity   92.9%;  Pred. No. 1.6e+05;
  Matches   13;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     11 taccgatgaactaa 24
          ||||||||| ||||
Db      4 TACCGATGAGCTAA 17


RESULT   4
AR152501
LOCUS       AR152501       20 bp    DNA             PAT       08-AUG-2001
DEFINITION  Sequence 52 from patent US 6235283.
ACCESSION   AR152501
VERSION     AR152501.1  GI:15120033
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 20)
  AUTHORS   Cobon,G.Stewart, Moore,J.Terry, Johnston,L.Anthony.Yorke,
            Willadsen,P., Kemp,D.Harold, Sriskantha,A., Riding,G.Alfred and
            Rand,K.Norman.
  TITLE     DNA encoding a cell membrane glycoprotein of a tick gut
  JOURNAL   Patent: US 6235283-A 52 22-MAY-2001;
FEATURES             Location/Qualifiers
     source          1. .20
                     /organism="unknown"
BASE COUNT        6 a      5 c      6 g      3 t
ORIGIN


  Query Match             51.7%;  Score 12.4;  DB 6;  Length 20;
  Best Local Similarity   92.9%;  Pred. No. 1.6e+05;
  Matches   13;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      3 gaacttggtaccga 16
          ||||| ||||||||
Db      4 GAACTGGGTACCGA 17


RESULT   5
I32378
LOCUS       I32378         20 bp    DNA             PAT       06-FEB-1997
DEFINITION  Sequence 52 from patent US 5587311.
ACCESSION   I32378
VERSION     I32378.1  GI:1823169
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 20)
  AUTHORS   Cobon,G.S., Moore,J.T., Johnston,L.A.Y., Willadsen,P., Kemp,D.H.,
            Sriskantha,A., Riding,G.A. and Rand,K.N.
  TITLE     DNA encoding a cell membrane glycoprotein of a tick gut
  JOURNAL   Patent: US 5587311-A 52 24-DEC-1996;
FEATURES             Location/Qualifiers
     source          1. .20
                     /organism="unknown"
BASE COUNT        6 a      5 c      6 g      3 t
ORIGIN


  Query Match             51.7%;  Score 12.4;  DB 6;  Length 20;
  Best Local Similarity   92.9%;  Pred. No. 1.6e+05;
  Matches   13;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      3 gaacttggtaccga 16
          ||||| ||||||||
Db      4 GAACTGGGTACCGA 17


RESULT   6
AR152503
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 6, 2002, 09:15:15 ; Search time 71.9 Seconds
(without alignments)
5.008 Million cell updates/sec

Title: US-09-897-042-18
Perfect score: 85
Sequence: 1 MNLVPMNPLVMNEFMN 16

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	42	49.4	775	1	US-07-603-133B-13	Sequence 13, Appl
2	40	47.1	775	1	US-07-603-133B-14	Sequence 14, Appl
3	40	47.1	776	1	US-07-603-133B-17	Sequence 17, Appl
4	40	47.1	776	1	US-07-603-133B-20	Sequence 20, Appl
5	40	47.1	776	3	US-08-089-397A-15	Sequence 15, Appl
6	40	47.1	908	2	US-08-249-380-2	Sequence 2, Appli
7	38	44.7	510	1	US-08-249-112-3	Sequence 3, Appli
8	38	44.7	510	5	PCT-US95-06556-3	Sequence 3, Appli
9	38	44.7	694	2	US-08-701-240-2	Sequence 2, Appli
10	38	44.7	694	4	US-09-138-236-2	Sequence 2, Appli
11	38	44.7	775	1	US-07-603-133B-12	Sequence 12, Appl
12	38	44.7	776	1	US-07-603-133B-18	Sequence 18, Appl
13	38	44.7	2182	2	US-08-487-826B-16	Sequence 16, Appl
14	37	43.5	195	1	US-08-063-552-9	Sequence 9, Appli
15	37	43.5	195	5	PCT-US93-05704-9	Sequence 9, Appli
16	37	43.5	695	2	US-08-701-240-4	Sequence 4, Appli
17	37	43.5	695	4	US-09-138-236-4	Sequence 4, Appli
18	37	43.5	1375	4	US-09-210-361-4	Sequence 4, Appli
19	37	43.5	1475	3	US-09-007-999-2	Sequence 2, Appli
20	37	43.5	1475	4	US-09-210-361-2	Sequence 2, Appli
21	36	42.4	409	2	US-09-031-485-15	Sequence 15, Appl
22	36	42.4	409	2	US-08-847-429A-15	Sequence 15, Appl
23	36	42.4	409	3	US-09-075-215A-17	Sequence 17, Appl
24	36	42.4	409	3	US-09-065-474-15	Sequence 15, Appl
25	36	42.4	712	2	US-08-468-576B-17	Sequence 17, Appl
26	36	42.4	712	2	US-08-468-579B-17	Sequence 17, Appl
27	36	42.4	712	3	US-08-468-577B-17	Sequence 17, Appl
28	36	42.4	1745	2	US-09-031-485-33	Sequence 33, Appl
29	36	42.4	1745	2	US-08-847-429A-33	Sequence 33, Appl
30	36	42.4	1745	3	US-09-065-474-33	Sequence 33, Appl
31	35.5	41.8	113	4	US-09-180-077-6	Sequence 6, Appli
32	35.5	41.8	113	4	US-09-180-077-11	Sequence 11, Appl
33	35	41.2	311	2	US-08-775-009-32	Sequence 32, Appl
34	35	41.2	311	2	US-08-775-009-33	Sequence 33, Appl
35	35	41.2	353	4	US-09-118-442-6	Sequence 6, Appli
36	35	41.2	353	4	US-09-677-064-6	Sequence 6, Appli
37	35	41.2	2296	2	US-08-286-819A-27	Sequence 27, Appl
38	35	41.2	2296	3	US-08-980-357-27	Sequence 27, Appl
39	34	40.0	339	1	US-08-626-994A-3	Sequence 3, Appli
40	34	40.0	339	3	US-08-957-742-3	Sequence 3, Appli
41	34	40.0	364	1	US-08-626-994A-1	Sequence 1, Appli
42	34	40.0	364	3	US-08-957-742-1	Sequence 1, Appli
43	34	40.0	375	2	US-08-446-875-8	Sequence 8, Appli
44	34	40.0	375	2	US-08-102-385G-8	Sequence 8, Appli
45	34	40.0	432	2	US-08-700-152A-4	Sequence 4, Appli
46	34	40.0	519	2	US-08-725-736D-2	Sequence 2, Appli
47	34	40.0	519	3	US-09-162-368B-2	Sequence 2, Appli
48	34	40.0	519	4	US-09-161-877B-2	Sequence 2, Appli
49	34	40.0	671	6	5266464-2	Patent No. 5266464
50	34	40.0	685	3	US-09-031-563-21	Sequence 21, Appl

ALIGNMENTS

```
RESULT   1
US-07-603-133B-13
; Sequence 13, Application US/07603133B
; Patent No. 5298244
;  GENERAL INFORMATION:
;    APPLICANT:  Redmond, Mark J.
;    APPLICANT:  Ijaz, Mohammed K.
;    APPLICANT:  Parker, Michael D.
;    TITLE OF INVENTION:  ASSEMBLED VIRAL PARTICLES AND THEIR
;    TITLE OF INVENTION:  USE IN A VACCINE TO ROTAVIRAL DISEASE
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Morrison & Foerster
;      STREET:  545 Middlefield Road, Suite 200
;      CITY:  Menlo Park
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94025
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/603,133B
;      FILING DATE:  19901025
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Robins, Roberta L.
;      REGISTRATION NUMBER:  33,208
;      REFERENCE/DOCKET NUMBER:  9313-0004.00
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 327-7250
;      TELEFAX:  (415) 327-2951
;      TELEX:  706141
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  775 amino acids
;      TYPE:  AMINO ACID
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-07-603-133B-13
```

```
  Query Match             49.4%;  Score 42;  DB 1;  Length 775;
  Best Local Similarity   43.8%;  Pred. No. 28;
  Matches    7;  Conservative    4;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 MNLVPMNPLVMNEFMN 16
          :||:  || |:  |:|
Db    741 LNLIKSNPNVLRNFIN 756


RESULT   2
US-07-603-133B-14
; Sequence 14, Application US/07603133B
; Patent No. 5298244
;  GENERAL INFORMATION:
;    APPLICANT:  Redmond, Mark J.
;    APPLICANT:  Ijaz, Mohammed K.
;    APPLICANT:  Parker, Michael D.
     TITLE OF INVENTION:  ASSEMBLED VIRAL PARTICLES AND THEIR
     TITLE OF INVENTION:  USE IN A VACCINE TO ROTAVIRAL DISEASE
     NUMBER OF SEQUENCES:  30
,    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Morrison & Foerster
;      STREET:  545 Middlefield Road, Suite 200
;      CITY:  Menlo Park
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94025
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/603,133B
;      FILING DATE:  19901025
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Robins, Roberta L.
;      REGISTRATION NUMBER:  33,208
;      REFERENCE/DOCKET NUMBER:  9313-0004.00
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 327-7250
;      TELEFAX:  (415) 327-2951
;      TELEX:  706141
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
       LENGTH:  775 amino acids
       TYPE:  AMINO ACID
,      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-07-603-133B-14


  Query Match             47.1%;  Score 40;  DB 1;  Length 775;
  Best Local Similarity   46.7%;  Pred. No. 60;
  Matches    7;  Conservative    3;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      2 NLVPMNPLVMNEFMN 16
          ||:  || |:  |:|
Db    742 NLIKSNPNVLRNFIN 756


RESULT   3
US-07-603-133B-17
; Sequence 17, Application US/07603133B
; Patent No. 5298244
;  GENERAL INFORMATION:
;    APPLICANT:  Redmond, Mark J.
;    APPLICANT:  Ijaz, Mohammed K.
;    APPLICANT:  Parker, Michael D.
;    TITLE OF INVENTION:  ASSEMBLED VIRAL PARTICLES AND THEIR
;    TITLE OF INVENTION:  USE IN A VACCINE TO ROTAVIRAL DISEASE
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Morrison & Foerster
;      STREET:  545 Middlefield Road, Suite 200
;      CITY:  Menlo Park
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94025
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/603,133B
;      FILING DATE:  19901025
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Robins, Roberta L.
;      REGISTRATION NUMBER:  33,208
;      REFERENCE/DOCKET NUMBER:  9313-0004.00
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 327-7250
;      TELEFAX:  (415) 327-2951
;      TELEX:  706141
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  776 amino acids
;      TYPE:  AMINO ACID
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-07-603-133B-17


  Query Match             47.1%;  Score 40;  DB 1;  Length 776;
  Best Local Similarity   43.8%;  Pred. No. 60;
  Matches    7;  Conservative    5;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 MNLVPMNPLVMNEFMN 16
          :||:  :| |: ||:|
Db    742 LNLLRSDPRVLREFIN 757


RESULT   4
US-07-603-133B-20
; Sequence 20, Application US/07603133B
; Patent No. 5298244
;  GENERAL INFORMATION:
;    APPLICANT:  Redmond, Mark J.
;    APPLICANT:  Ijaz, Mohammed K.
;    APPLICANT:  Parker, Michael D.
;    TITLE OF INVENTION:  ASSEMBLED VIRAL PARTICLES AND THEIR
;    TITLE OF INVENTION:  USE IN A VACCINE TO ROTAVIRAL DISEASE
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Morrison & Foerster
;      STREET:  545 Middlefield Road, Suite 200
;      CITY:  Menlo Park
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94025
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/603,133B
;      FILING DATE:  19901025
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Robins, Roberta L.
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 6, 2002, 09:18:43 ; Search time 103.34 Seconds
(without alignments)
11.469 Million cell updates/sec

Title: US-09-897-042-18
Perfect score: 85
Sequence: 1 MNLVPMNPLVMNEFMN 16

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : A_Geneseq_1101:*
1: /SIDS2/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	40	47.1	776	8	AAP71568	Sequence of rotavi
2	40	47.1	776	13	AAR24293	VP4 protein of str
3	40	47.1	776	21	AAB07418	Amino acid sequenc
4	40	47.1	906	14	AAR30053	E. coli heat shock
5	40	47.1	908	14	AAR33429	E. coli heat shock
6	39	45.9	107	21	AAG58560	Arabidopsis thalia
7	39	45.9	232	21	AAY59420	Catalpa inositol 1
8	39	45.9	427	21	AAG29799	Arabidopsis thalia
9	39	45.9	428	21	AAG29798	Arabidopsis thalia
10	39	45.9	443	21	AAG29797	Arabidopsis thalia
11	39	45.9	1997	21	AAB18287	Plasmodium falcipa
12	39	45.9	2001	22	AAB20062	Arabidopsis thalia
13	39	45.9	2001	22	AAB20063	Arabidopsis thalia
14	38	44.7	275	22	AAB95047	Human protein sequ
15	38	44.7	379	21	AAB11504	SEN virus protein
16	38	44.7	510	17	AAR88360	Caenorhabditis ele
17	38	44.7	522	22	AAG84964	Shrimp white spot
18	38	44.7	597	21	AAY58114	Haemorrhagic enter
19	38	44.7	694	18	AAW13097	Human glial tumour
20	38	44.7	776	10	AAP91222	Rhesus rotavirus g
21	38	44.7	1091	16	AAR70982	Megakaryocyte stem
22	38	44.7	1091	16	AAR70983	Megakaryocyte stem
23	38	44.7	2182	18	AAW22476	Plasmodium var-1.
24	38	44.7	2182	21	AAY77906	Plasmodium var-1 p
25	37	43.5	195	15	AAR47339	Peptide fragment o
26	37	43.5	279	21	AAB42870	Human ORFX ORF2634
27	37	43.5	586	22	AAM38947	Human polypeptide
28	37	43.5	695	18	AAW13098	Mouse glial tumour
29	37	43.5	761	20	AAW99084	Non-B, non-C, non-
30	37	43.5	1330	21	AAY81947	Heterosigma akashi
31	36	42.4	151	19	AAW52130	Putative insectici
32	36	42.4	201	22	AAM17341	Peptide #3775 enco
33	36	42.4	201	22	AAM29848	Peptide #3885 enco
34	36	42.4	201	22	AAM05030	Peptide #3712 enco
35	36	42.4	219	22	AAB63630	Human gastric canc
36	36	42.4	244	20	AAY15237	ULBP-1 amino acid
37	36	42.4	252	21	AAY28422	Soybean branched c
38	36	42.4	268	22	AAB63629	Human gastric canc
39	36	42.4	270	20	AAY15241	ULBP1-LZ fusion po
40	36	42.4	278	21	AAG36698	Arabidopsis thalia
41	36	42.4	278	21	AAG37341	Arabidopsis thalia
42	36	42.4	278	21	AAG44322	Arabidopsis thalia
43	36	42.4	356	21	AAY74607	Neisseria meningit
44	36	42.4	363	21	AAY74606	Neisseria gonorrhe
45	36	42.4	366	22	AAG91960	C glutamicum prote
46	36	42.4	372	21	AAG36697	Arabidopsis thalia
47	36	42.4	372	21	AAG37340	Arabidopsis thalia
48	36	42.4	372	21	AAG44321	Arabidopsis thalia
49	36	42.4	389	21	AAG37339	Arabidopsis thalia
50	36	42.4	389	21	AAG44320	Arabidopsis thalia

ALIGNMENTS

RESULT 1
AAP71568
ID AAP71568 standard; Protein; 776 AA.
XX
AC AAP71568;
XX
DT 24-MAY-1991 (first entry)
XX
DE Sequence of rotavirus outer shell protein VP3 serotype C486 bovine.
XX
KW Diarrhoea; gastrointestinal disorder; RNA virus; vaccine.
XX
OS Rotavirus.
XX
FH Key Location/Qualifiers
FT Region 232..254
FT /note= "A fragment of VP3 with this SQ is claimed"
XX
PN AU8666987-A.
XX
PD 02-JUL-1987.
XX
PF 23-DEC-1986; 86AU-0017981.
XX
PR 03-SEP-1986; 86US-0903325.
PR 26-DEC-1985; 85US-0813661.
XX
PA (UYSA-) UNIV OF SASKATCHEWA.
XX

PI Sabara MIJ, Frenchick PJ, Potter AA, Ijaz MK, Gilchrist JE;
XX
DR WPI; 1987-228567/33.
XX
PT New peptide fragments of rota:viral proteins - useful conjugates
PT in vaccines for protecting against gastrointestinal disorders and
PT diarrhoea
XX
PS Disclosure; Fig 3; 84pp; English.
XX
CC The peptide fragments of glycoprotein VP7 and proteins VP6 and VP3
CC of rotaviruses are useful when attached to carriers as vaccines for
CC birds and mammals, including man. The vaccines confer protection
CC against gastrointestinal disorders and diarrhoea produced by the
CC rotaviruses. For use in vaccines the peptides are covalently linked
CC to eg, keyhole limpet haemocyanin, BSA, ovalbumin, poly-L-lysine,
CC or VP6 bovine rotavirus protein. An adjuvant may be included.
XX
SQ Sequence 776 AA;

Query Match 47.1%; Score 40; DB 8; Length 776;
Best Local Similarity 43.8%; Pred. No. 99;
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

Qy 1 MNLVPMNPLVMNEFMN 16
:||: :| |: ||:|
Db 742 lnllrsdprvlrefin 757

RESULT 2
AAR24293
ID AAR24293 standard; Protein; 776 AA.
XX
AC AAR24293;
XX
DT 18-NOV-1992 (first entry)
XX
DE VP4 protein of strain C486 (bovine).
XX
KW Vaccine; rotavirus; viral protein; virus; capsid;
KW VP6; VP4; VP7.
XX
OS Bovine C486 rotavirus.
XX
PN WO9207941-A.
XX
PD 14-MAY-1992.
XX
PF 24-OCT-1991; 91WO-CA00376.
XX
PR 25-OCT-1990; 90US-0603133.
XX
PA (UYSA-) UNIV SASKATCHEWAN.
XX
PI Ijaz MK, Parker MD, Redmond MJ;
XX
DR WPI; 1992-183676/22.
DR N-PSDB; AAQ25167.
XX
PT Assembled viral particles useful as a rota-viral vaccine -
PT contain the inner capsid protein VP6 in combination with either
PT or both of the outer capsid proteins VP4 and VP7
XX
PS Disclosure; Fig 4(1-9); 103pp; English.
XX
CC Assembled viral particles include the inner capsid protein, VP6,
CC in combination with either or both of the outer capsid proteins,
CC VP4 and VP7. These assemblies can be used in vaccine compsns.
CC for the treatment and prevention of rotaviral disease.
XX
SQ Sequence 776 AA;

Query Match 47.1%; Score 40; DB 13; Length 776;
Best Local Similarity 43.8%; Pred. No. 99;
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

Qy 1 MNLVPMNPLVMNEFMN 16
:||: :| |: ||:|
Db 742 lnllrsdprvlrefin 757

RESULT 3
AAB07418
ID AAB07418 standard; protein; 776 AA.
XX
AC AAB07418;
XX
DT 20-OCT-2000 (first entry)
XX
DE Amino acid sequence of a viral protein 7 (VP7) of rotavirus.
XX
KW Immunochemical; rotavirus; VP6; viral protein; VP4; antibody response;
KW vaccine; rotavirus infection.
XX
OS Rotavirus.
XX
PN US6086880-A.
XX
PD 11-JUL-2000.
XX
PF 07-JUL-1993; 93US-0089397.
XX
PR 03-SEP-1986; 86US-0903325.
PR 07-SEP-1988; 88US-0241761.
PR 27-FEB-1991; 91US-0661859.
PR 26-DEC-1985; 85US-0813661.
PR 12-JUL-1990; 90US-0552350.
PR 10-DEC-1990; 90US-0626041.
XX
PA (UYSA-) UNIV SASKATCHEWAN.
XX
PI Sabara MIJ, Frenchick PJ, Potter AA, Gilchrist JE, Redmond MJ;
PI Ijaz MK;
XX
DR WPI; 2000-498191/44.
XX
PT Peptide useful as vaccine for protection against infection by rotavirus
PT comprises specified subunits of rotavirus VP6 and VP4 viral proteins -
XX
PS Disclosure; Fig 1; 34pp; English.
XX
CC The specification describes peptides which are useful as
CC immunochemicals. The peptides comprise subunits 40-60 of rotavirus VP6
CC viral protein, 232-255 of VP4 (both optionally modified to facilitate
CC binding to carrier), where one or more of valine residues of VP4 is
CC substituted at positions 234, 236 or 255 by alanine, or subunits 240-248
CC of VP4 which is optionally modified to facilitate covalent coupling to a
CC carrier other than VP6. Rotavirus VP4 subunit was found to be capable of
CC plaque reduction when mixed with infection virus. The peptides are
CC useful for initiating an antibody response in a mammal against
CC rotaviruses. The peptides are also useful as vaccines to protect against
CC rotavirus infection and as diagnostic tools to detect the presence of
CC rotaviral infection. VP4 subunits are useful for prophylactic protection
CC or for therapy with respect to rotaviral infection. AAB07553-55 and
CC AAB07418 represent rotavirus VP7 proteins.
XX
SQ Sequence 776 AA;

Query Match 47.1%; Score 40; DB 21; Length 776;
Best Local Similarity 43.8%; Pred. No. 99;
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

Qy 1 MNLVPMNPLVMNEFMN 16

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 6, 2002, 09:33:20 ; Search time 135.97 Seconds
(without alignments)
17.212 Million cell updates/sec

Title: US-09-897-042-18
Perfect score: 85
Sequence: 1 MNLVPMNPLVMNEFMN 16

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

imum DB seq length: 0
imum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : SPTREMBL_17:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

lt No.	Score	% Query Match	Length	DB	ID	Description
1	44	51.8	775	12	Q9WAL4	Q9wal4 human rotav
2	44	51.8	775	12	Q9WAL1	Q9wal1 human rotav
3	44	51.8	775	12	Q9WAL0	Q9wal0 human rotav
4	44	51.8	775	12	Q9WAK8	Q9wak8 human rotav
5	44	51.8	775	12	Q9WAK4	Q9wak4 human rotav
6	44	51.8	775	12	Q9WAK3	Q9wak3 human rotav
7	44	51.8	775	12	Q9WAK2	Q9wak2 human rotav
8	44	51.8	775	12	Q9WAK0	Q9wak0 human rotav
9	44	51.8	775	12	Q9W8J0	Q9w8j0 human rotav
10	43	50.6	179	8	Q9GAG5	Q9gag5 polyrhachis
11	43	50.6	500	10	Q9S9W4	Q9s9w4 arabidopsis
12	42	49.4	226	8	Q9G2G2	Q9g2g2 phytomyza n
13	42	49.4	319	8	Q9B6G2	Q9b6g2 thryonomys
14	42	49.4	330	6	Q9BFF1	Q9bff1 sylvilagus
15	42	49.4	330	6	Q9BFF0	Q9bff0 ochotona hy
16	42	49.4	346	2	Q9RVT6	Q9rvt6 deinococcus
17	42	49.4	775	12	Q82012	Q82012 human rotav
18	42	49.4	775	12	Q82118	Q82118 human rotav
19	42	49.4	775	12	Q86111	Q86111 rotavirus g
20	42	49.4	775	12	Q86185	Q86185 rotavirus s
21	42	49.4	775	12	Q9WAL7	Q9wal7 human rotav
22	42	49.4	775	12	Q9WAL6	Q9wal6 human rotav
23	42	49.4	775	12	Q9WAL3	Q9wal3 human rotav
24	42	49.4	775	12	Q9WAK9	Q9wak9 human rotav
25	42	49.4	775	12	Q9WAK5	Q9wak5 human rotav
26	42	49.4	775	12	Q9WAK1	Q9wak1 human rotav
27	42	49.4	775	12	Q9E3V4	Q9e3v4 human rotav
28	42	49.4	775	12	Q9DUU6	Q9duu6 human rotav
29	42	49.4	775	12	Q99IN3	Q99in3 human rotav
30	42	49.4	775	12	Q99IN2	Q99in2 human rotav
31	42	49.4	775	12	Q99IN1	Q99in1 human rotav
32	42	49.4	775	12	Q99IN0	Q99in0 human rotav
33	42	49.4	775	12	Q99IM9	Q99im9 human rotav
34	42	49.4	775	12	Q99IM8	Q99im8 human rotav
35	42	49.4	775	12	Q99IM7	Q99im7 human rotav
36	42	49.4	775	12	Q98VL3	Q98vl3 human rotav
37	41	48.2	179	8	Q9GAH8	Q9gah8 oecophylla
38	41	48.2	184	8	O03164	O03164 greya solen
39	41	48.2	185	8	Q36040	Q36040 tridentafor
40	41	48.2	576	2	Q9CN68	Q9cn68 pasteurella
41	41	48.2	620	5	O16519	O16519 caenorhabdi
42	41	48.2	994	5	O77070	O77070 placopecten
43	40	47.1	100	8	Q9G2E1	Q9g2e1 hylomys sui
44	40	47.1	145	8	Q9G7P3	Q9g7p3 diasemopsis
45	40	47.1	153	10	Q9M2J3	Q9m2j3 arabidopsis
46	40	47.1	159	8	Q34592	Q34592 greya mitel
47	40	47.1	170	8	Q34614	Q34614 greya obscu
48	40	47.1	174	8	Q9TBZ9	Q9tbz9 greya solen
49	40	47.1	179	8	Q9GAG8	Q9gag8 polyrhachis
50	40	47.1	180	8	Q34562	Q34562 greya enchr

ALIGNMENTS

RESULT 1
Q9WAL4
ID Q9WAL4 PRELIMINARY; PRT; 775 AA.
AC Q9WAL4;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE VP4.
OS Human rotavirus 3.
OC Viruses; dsRNA viruses; Reoviridae; Rotavirus.
OX NCBI_TaxID=36430;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TK-28;
RX MEDLINE=99244278; PubMed=10229272;
RA Cao X.R., Akihara S., Fang Z.Y., Nakagomi O., Ushijima H.;
RT "Genetic variation in the VP4 and NSP4 genes of human rotavirus
RT serotype 3 (G3 type) isolated in China and Japan.";
RL Microbiol. Immunol. 43:171-175(1999).
DR EMBL; AB008288; BAA77553.1; -.
DR InterPro; IPR000416; Cap_VP4.
DR Pfam; PF00426; VP4; 1.
SQ SEQUENCE 775 AA; 87276 MW; B4D8D22DFD7CCED6 CRC64;

Query Match 51.8%; Score 44; DB 12; Length 775;
Best Local Similarity 43.8%; Pred. No. 32;
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

```
Qy       1 MNLVPMNPLVMNEFMN 16
           :||:  || |: :|:|
Db     741 LNLIKSNPNVLRDFIN 756
```

RESULT 2
Q9WAL1
ID Q9WAL1 PRELIMINARY; PRT; 775 AA.

```
AC   Q9WAL1;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   VP4.
OS   Human rotavirus 3.
OC   Viruses; dsRNA viruses; Reoviridae; Rotavirus.
OX   NCBI_TaxID=36430;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AI-75;
RX   MEDLINE=99244278; PubMed=10229272;
RA   Cao X.R., Akihara S., Fang Z.Y., Nakagomi O., Ushijima H.;
RT   "Genetic variation in the VP4 and NSP4 genes of human rotavirus
RT   serotype 3 (G3 type) isolated in China and Japan.";
RL   Microbiol. Immunol. 43:171-175(1999).
DR   EMBL; AB008285; BAA77550.1; -.
DR   InterPro; IPR000416; Cap_VP4.
DR   Pfam; PF00426; VP4; 1.
SQ   SEQUENCE   775 AA;  87408 MW;  AA1F451628B4F9EB CRC64;


  Query Match             51.8%;  Score 44;  DB 12;  Length 775;
  Best Local Similarity   43.8%;  Pred. No. 32;
  Matches    7;  Conservative    5;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 MNLVPMNPLVMNEFMN 16
           :||:  || |: :|:|
Db     741 LNLIKSNPNVLRDFIN 756


RESULT   3
Q9WAL0
ID   Q9WAL0      PRELIMINARY;      PRT;   775 AA.
AC   Q9WAL0;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   VP4.
OS   Human rotavirus 3.
OC   Viruses; dsRNA viruses; Reoviridae; Rotavirus.
OX   NCBI_TaxID=36430;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AI-53;
RX   MEDLINE=99244278; PubMed=10229272;
RA   Cao X.R., Akihara S., Fang Z.Y., Nakagomi O., Ushijima H.;
RT   "Genetic variation in the VP4 and NSP4 genes of human rotavirus
RT   serotype 3 (G3 type) isolated in China and Japan.";
RL   Microbiol. Immunol. 43:171-175(1999).
DR   EMBL; AB008284; BAA77549.1; -.
DR   InterPro; IPR000416; Cap_VP4.
DR   Pfam; PF00426; VP4; 1.
SQ   SEQUENCE   775 AA;  87322 MW;  F9C2C714320A183D CRC64;


  Query Match             51.8%;  Score 44;  DB 12;  Length 775;
  Best Local Similarity   43.8%;  Pred. No. 32;
  Matches    7;  Conservative    5;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 MNLVPMNPLVMNEFMN 16
           :||:  || |: :|:|
Db     741 LNLIKSNPNVLRDFIN 756


RESULT   4
Q9WAK8
ID   Q9WAK8      PRELIMINARY;      PRT;   775 AA.
AC   Q9WAK8;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   VP4.
OS   Human rotavirus 3.
OC   Viruses; dsRNA viruses; Reoviridae; Rotavirus.
OX   NCBI_TaxID=36430;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AI-36;
RX   MEDLINE=99244278; PubMed=10229272;
RA   Cao X.R., Akihara S., Fang Z.Y., Nakagomi O., Ushijima H.;
RT   "Genetic variation in the VP4 and NSP4 genes of human rotavirus
RT   serotype 3 (G3 type) isolated in China and Japan.";
RL   Microbiol. Immunol. 43:171-175(1999).
DR   EMBL; AB008282; BAA77547.1; -.
DR   InterPro; IPR000416; Cap_VP4.
DR   Pfam; PF00426; VP4; 1.
SQ   SEQUENCE   775 AA;  87309 MW;  9FBB7983A8BC5B38 CRC64;


  Query Match             51.8%;  Score 44;  DB 12;  Length 775;
  Best Local Similarity   43.8%;  Pred. No. 32;
  Matches    7;  Conservative    5;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 MNLVPMNPLVMNEFMN 16
           :||:  || |: :|:|
Db     741 LNLIKSNPNVLRDFIN 756


RESULT   5
Q9WAK4
ID   Q9WAK4      PRELIMINARY;      PRT;   775 AA.
AC   Q9WAK4;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   VP4.
OS   Human rotavirus 3.
OC   Viruses; dsRNA viruses; Reoviridae; Rotavirus.
OX   NCBI_TaxID=36430;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MO;
RX   MEDLINE=99244278; PubMed=10229272;
RA   Cao X.R., Akihara S., Fang Z.Y., Nakagomi O., Ushijima H.;
RT   "Genetic variation in the VP4 and NSP4 genes of human rotavirus
RT   serotype 3 (G3 type) isolated in China and Japan.";
RL   Microbiol. Immunol. 43:171-175(1999).
DR   EMBL; AB008278; BAA77543.1; -.
DR   InterPro; IPR000416; Cap_VP4.
DR   Pfam; PF00426; VP4; 1.
SQ   SEQUENCE   775 AA;  87288 MW;  E4CE723A8124818F CRC64;


  Query Match             51.8%;  Score 44;  DB 12;  Length 775;
  Best Local Similarity   43.8%;  Pred. No. 32;
  Matches    7;  Conservative    5;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 MNLVPMNPLVMNEFMN 16
           :||:  || |: :|:|
Db     741 LNLIKSNPNVLRDFIN 756


RESULT   6
Q9WAK3
ID   Q9WAK3      PRELIMINARY;      PRT;   775 AA.
AC   Q9WAK3;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   VP4.
OS   Human rotavirus 3.
OC   Viruses; dsRNA viruses; Reoviridae; Rotavirus.
OX   NCBI_TaxID=36430;
RN   [1]
RP   SEQUENCE FROM N.A.
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 6, 2002, 09:34:42 ; Search time 68.88 Seconds
(without alignments)
8.517 Million cell updates/sec

Title: US-09-897-042-18
Perfect score: 85
Sequence: 1 MNLVPMNPLVMNEFMN 16

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	43	50.6	2054	1	YCF2_PINTH	P41653 pinus thunb
2	42	49.4	775	1	VP4_ROTHK	P13842 human rotav
3	42	49.4	775	1	VP4_ROTHL	P21284 human rotav
4	41	48.2	241	1	YF45_METJA	Q58940 methanococc
5	41	48.2	576	1	CYDC_HAEIN	P45081 haemophilus
6	40	47.1	184	1	21KD_ZYMMO	P15256 zymomonas m
7	40	47.1	334	1	Y009_BORBU	O51042 borrelia bu
8	40	47.1	775	1	VP4_NCDV	P17465 nebraska ca
9	40	47.1	775	1	VP4_ROTH5	P30214 human rotav
10	40	47.1	775	1	VP4_ROTHD	P11196 human rotav
11	40	47.1	775	1	VP4_ROTHP	P11195 human rotav
12	40	47.1	775	1	VP4_ROTHW	P11193 human rotav
13	40	47.1	776	1	VP41_ROTS1	P04508 simian 11 r
14	40	47.1	776	1	VP4_ROTB5	P36305 bovine rota
15	40	47.1	776	1	VP4_ROTBC	P08713 bovine rota
16	40	47.1	776	1	VP4_ROTSF	P17463 simian 11 r
17	40	47.1	908	1	H104_YEAST	P31539 saccharomyc
18	39	45.9	536	1	CCA4_DROME	Q9ve00 drosophila
19	39	45.9	590	1	NU5M_DINSE	O79556 dinodon sem
20	39	45.9	775	1	VP4_ROTHV	P11194 human rotav
21	39	45.9	776	1	VP4_ROTHQ	Q08778 human rotav
22	39	45.9	1018	1	VGNM_BPMV	P23009 bean-pod mo
23	38	44.7	200	1	ATKC_ANASL	Q9r6w9 anabaena sp
24	38	44.7	314	1	MHPB_ECOLI	P54711 escherichia
25	38	44.7	348	1	OPSD_GLOME	O62792 globicephal
26	38	44.7	413	1	EF1A_HELVI	P55276 heliothis v
27	38	44.7	463	1	EF1A_BOMMO	P29520 bombyx mori
28	38	44.7	775	1	VP4_ROTF1	P39034 feline rota
29	38	44.7	775	1	VP4_ROTH3	P39033 human rotav
30	38	44.7	775	1	VP4_ROTHJ	Q01641 human rotav
31	38	44.7	776	1	VP42_ROTS1	P12976 simian 11 r
32	38	44.7	776	1	VP4_ROTD9	Q06894 canine rota
33	38	44.7	776	1	VP4_ROTEH	Q02945 equine rota
34	38	44.7	776	1	VP4_ROTF6	Q07416 feline rota
35	38	44.7	776	1	VP4_ROTHY	Q06895 human rotav
36	38	44.7	776	1	VP4_ROTRH	P12473 rhesus rota
37	38	44.7	776	1	VP4_ROTSS	P17464 simian 11 r
38	38	44.7	1142	1	SPK_HUMAN	Q92797 homo sapien
39	38	44.7	2179	1	POLG_EC23W	O73556 e genome po
40	37	43.5	284	1	SIC1_YEAST	P38634 saccharomyc
41	37	43.5	347	1	NU2M_PIG	O79875 sus scrofa
42	37	43.5	379	1	CYB_LOCMI	Q36427 locusta mig
43	37	43.5	389	1	BMR1_BACSU	P33449 bacillus su
44	37	43.5	439	1	EF12_EUPCR	Q27140 euplotes cr
45	37	43.5	447	1	EF12_DAUCA	P34823 daucus caro
46	37	43.5	463	1	EF11_DROME	P08736 drosophila
47	37	43.5	536	1	YC42_SYNY3	P42349 synechocyst
48	37	43.5	712	1	RR3_CHLRE	Q08365 chlamydomon
49	37	43.5	729	1	KAR3_YEAST	P17119 saccharomyc
50	37	43.5	765	1	TRPG_PHYBL	P20409 p anthranil

ALIGNMENTS

RESULT 1
YCF2_PINTH
ID YCF2_PINTH STANDARD; PRT; 2054 AA.
AC P41653;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL 244.6 KDA PROTEIN YCF2 (ORF 2054).
GN YCF2.
OS Pinus thunbergii (Green pine) (Japanese black pine).
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.
OX NCBI_TaxID=3350;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95024047; PubMed=7937893;
RA Wakasugi T., Tsudzuki J., Ito S., Nakashima K., Tsudzuki T.,
RA Sugiura M.;
RT "Loss of all ndh genes as determined by sequencing the entire
RT chloroplast genome of the black pine Pinus thunbergii.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:9794-9798(1994).
CC -!- FUNCTION: NOT YET KNOWN.
CC -!- SIMILARITY: BELONGS TO THE YCF2 FAMILY.
CC ----------

CC ----------
DR EMBL; D17510; BAA04460.1; -.
DR InterPro; IPR001939; AAA_subfam.
DR Pfam; PF00004; AAA; 1.
KW Chloroplast; Hypothetical protein.
SQ SEQUENCE 2054 AA; 244604 MW; 6F5E92D078E33A9A CRC64;

Query Match 50.6%; Score 43; DB 1; Length 2054;
Best Local Similarity 46.7%; Pred. No. 34;
Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

```
Qy       2 NLVPMNPLVMNEFMN 16
           | :|:|||:: | |
Db     717 NKIPINPLIIDFFDN 731
```

RESULT 2
VP4_ROTHK

```
ID   VP4_ROTHK      STANDARD;      PRT;   775 AA.
AC   P13842;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   OUTER CAPSID PROTEIN VP4 (HEMAGGLUTININ) (OUTER LAYER PROTEIN VP4)
DE   [CONTAINS: OUTER CAPSID PROTEINS VP5 AND VP8].
GN   S4.
OS   Human rotavirus (strain KU).
OC   Viruses; dsRNA viruses; Reoviridae; Rotavirus.
OX   NCBI_TaxID=10952;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88230603; PubMed=2453680;
RA   Taniguchi K., Maloy W.L., Nishikawa K., Green K.Y., Hoshino Y.,
RA   Urasawa S., Kapikian A.Z., Chanock R.M., Gorziglia M.;
RT   "Identification of cross-reactive and serotype 2-specific
RT   neutralization epitopes on VP3 of human rotavirus.";
RL   J. Virol. 62:2421-2426(1988).
CC   -!- SUBCELLULAR LOCATION: OUTER CAPSID.
CC   -!- PTM: VP8 IS ONE OF TWO TRYPSIN CLEAVAGE PRODUCTS OF VP4; THE
CC       OTHER PRODUCT IS VP5.
CC   -!- SIMILARITY: BELONGS TO THE ROTAVIRUS VP4 PROTEIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M21014; AAA47334.1; -.
DR   PIR; A28844; VPXRHK.
DR   InterPro; IPR000416; Cap_VP4.
DR   Pfam; PF00426; VP4; 1.
KW   Coat protein; Glycoprotein.
FT   CHAIN         1    775       OUTER CAPSID PROTEIN VP4.
FT   CHAIN         1    240       OUTER CAPSID PROTEIN VP8.
FT   CHAIN       247    775       OUTER CAPSID PROTEIN VP5.
FT   CARBOHYD     32     32       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     56     56       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     97     97       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    132    132       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    324    324       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    583    583       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    589    589       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    592    592       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    599    599       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   775 AA;  87538 MW;  3222AE1997D0FDDF CRC64;

  Query Match             49.4%;  Score 42;  DB 1;  Length 775;
  Best Local Similarity   43.8%;  Pred. No. 17;
  Matches    7;  Conservative    4;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 MNLVPMNPLVMNEFMN 16
           :||:  || |:  |:|
Db     741 LNLIKSNPNVLRNFIN 756


RESULT   3
VP4_ROTHL
ID   VP4_ROTHL      STANDARD;      PRT;   775 AA.
AC   P21284;
DT   01-MAY-1991 (Rel. 18, Created)
DT   01-MAY-1991 (Rel. 18, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   OUTER CAPSID PROTEIN VP4 (HEMAGGLUTININ) (OUTER LAYER PROTEIN VP4)
DE   [CONTAINS: OUTER CAPSID PROTEINS VP5 AND VP8].
GN   S4.
OS   Human rotavirus (strain L26).
OC   Viruses; dsRNA viruses; Reoviridae; Rotavirus.
OX   NCBI_TaxID=10953;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91012813; PubMed=2170690;
RA   Taniguchi K., Urasawa T., Kobayashi N., Gorziglia M., Urasawa S.;
RT   "Nucleotide sequence of VP4 and VP7 genes of human rotaviruses with
RT   subgroup I specificity and long RNA pattern: implication for new G
RT   serotype specificity.";
RL   J. Virol. 64:5640-5644(1990).
CC   -!- SUBCELLULAR LOCATION: OUTER CAPSID.
CC   -!- PTM: VP8 IS ONE OF TWO TRYPSIN CLEAVAGE PRODUCTS OF VP4; THE
CC       OTHER PRODUCT IS VP5.
CC   -!- SIMILARITY: BELONGS TO THE ROTAVIRUS VP4 PROTEIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M58292; AAA47335.1; -.
DR   PIR; A36410; VPXRWL.
DR   PIR; C36410; VPXRWM.
DR   InterPro; IPR000416; Cap_VP4.
DR   Pfam; PF00426; VP4; 1.
KW   Coat protein; Glycoprotein.
FT   CHAIN         1    775       OUTER CAPSID PROTEIN VP4.
FT   CHAIN         1    240       OUTER CAPSID PROTEIN VP8.
FT   CHAIN       247    775       OUTER CAPSID PROTEIN VP5.
FT   CARBOHYD     32     32       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     56     56       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     97     97       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    132    132       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    150    150       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    195    195       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    324    324       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    583    583       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    589    589       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    599    599       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARIANT      51     51       R -> G (IN STRAIN L27).
FT   VARIANT     392    392       E -> K (IN STRAIN L27).
FT   VARIANT     405    405       F -> C (IN STRAIN L27).
SQ   SEQUENCE   775 AA;  87650 MW;  40CA498305DECD63 CRC64;


  Query Match             49.4%;  Score 42;  DB 1;  Length 775;
  Best Local Similarity   43.8%;  Pred. No. 17;
  Matches    7;  Conservative    4;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 MNLVPMNPLVMNEFMN 16
           :||:  || |:  |:|
Db     741 LNLIKSNPNVLRNFIN 756


RESULT   4
YF45_METJA
ID   YF45_METJA     STANDARD;      PRT;   241 AA.
AC   Q58940;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   HYPOTHETICAL PROTEIN MJ1545.
GN   MJ1545.
OS   Methanococcus jannaschii.
OC   Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC   Methanococcus.
OX   NCBI_TaxID=2190;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX   MEDLINE=96337999; PubMed=8688087;
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 6, 2002, 09:16:49 ; Search time 77.97 Seconds
(without alignments)
15.632 Million cell updates/sec

Title: US-09-897-042-18
Perfect score: 85
Sequence: 1 MNLVPMNPLVMNEFMN 16

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	43	50.6	500	2	G85069	hypothetical prote
2	43	50.6	2054	2	T07584	hypothetical prote
3	42	49.4	346	2	F75457	conserved hypothet
4	42	49.4	775	1	VPXRHK	outer layer protei
5	42	49.4	775	1	VPXRWL	outer layer protei
6	42	49.4	775	1	VPXRWM	outer layer protei
7	42	49.4	775	2	S52165	outer capsid prote
8	42	49.4	776	2	A48480	outer capsid prote
9	41	48.2	241	2	H64492	hypothetical prote
10	41	48.2	576	2	E64186	probable ATP-bindi
11	41	48.2	620	2	T31906	hypothetical prote
12	40	47.1	153	2	T45983	hypothetical prote
13	40	47.1	184	2	S06695	hypothetical prote
14	40	47.1	204	2	T03962	r40g3 protein - ri
15	40	47.1	334	2	A70101	hypothetical prote
16	40	47.1	362	2	T33502	hypothetical prote
17	40	47.1	382	2	E85082	hypothetical prote
18	40	47.1	382	2	T14186	hypothetical prote
19	40	47.1	656	2	T52064	dnaJ-like protein
20	40	47.1	775	1	VPXRT2	outer layer protei
21	40	47.1	775	1	VPXRW3	outer layer protei
22	40	47.1	775	1	VPXRW4	outer layer protei
23	40	47.1	775	1	VPXRWF	outer layer protei
24	40	47.1	775	1	VPXRW5	outer layer protei
25	40	47.1	776	1	VPXRT1	outer layer protei
26	40	47.1	776	1	VPXRB3	outer layer protei
27	40	47.1	776	1	JQ2022	outer layer protei
28	40	47.1	776	2	S24410	hypothetical outer
29	40	47.1	908	1	S61476	endopeptidase Clp
30	39.5	46.5	486	2	S31805	VP5 protein - porc
31	39.5	46.5	2910	2	T28156	DNA-directed RNA p
32	39	45.9	333	2	E86444	hypothetical prote
33	39	45.9	334	2	T46238	hypothetical prote
34	39	45.9	343	2	T33945	hypothetical prote
35	39	45.9	427	2	A84820	hypothetical prote
36	39	45.9	461	2	T51991	translation elonga
37	39	45.9	590	2	T11098	NADH dehydrogenase
38	39	45.9	775	1	VPXRW6	outer layer protei
39	39	45.9	799	2	F83549	probable ATP-depen
40	39	45.9	1018	1	GNWXG7	genome polyprotein
41	39	45.9	1997	2	F71607	DNA helicase II BR
42	39	45.9	2254	2	D86215	protein T6D22.14 [
43	38.5	45.3	419	2	T23949	hypothetical prote
44	38.5	45.3	883	2	T23948	hypothetical prote
45	38	44.7	63	2	T12121	NADH dehydrogenase
46	38	44.7	141	2	S08523	hypothetical prote
47	38	44.7	200	2	T46848	K+-transporting AT
48	38	44.7	314	2	D64762	2,3-dihydroxypheny
49	38	44.7	314	2	G85529	2,3-dihydroxypheny
50	38	44.7	343	2	G72218	conserved hypothet

ALIGNMENTS

RESULT 1
G85069
hypothetical protein AT4g05550 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001
C;Accession: G85069
R;anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Sp
Nature 402, 769-777, 1999
A;Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A;Reference number: A85001; MUID:20083488
A;Accession: G85069
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-500 <STO>
A;Cross-references: GB:NC_001268; NID:g7267315; PIDN:CAB81097.1; GSPDB:GN00140
C;Genetics:
A;Gene: AT4g05550
A;Map position: 4

Query Match 50.6%; Score 43; DB 2; Length 500;
Best Local Similarity 50.0%; Pred. No. 15;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

```
Qy      1 MNLVPMNPLVMNEFMN 16
          |||  :  :|: ||||
Db      1 MNLSSLEDIVLKEFMN 16
```

RESULT 2
T07584
hypothetical protein 2054 - Japanese black pine chloroplast
C;Species: chloroplast Pinus thunbergiana (Japanese black pine)
C;Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 18-Aug-2000
C;Accession: T07584
R;Wakasugi, T.; Tsudzuki, J.; Ito, S.; Nakashima, K.; Tsudzuki, T.; Sugiura, M.
Proc. Natl. Acad. Sci. U.S.A. 91, 9794-9798, 1994
A;Title: Loss of all ndh genes as determined by sequencing the entire chloroplast gen
A;Reference number: Z16030; MUID:95024047
A;Accession: T07584
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2054 <WAK>
A;Cross-references: EMBL:D17510; NID:g529643; PIDN:BAA04460.1; PID:g1262745
C;Genetics:
A;Genome: chloroplast
C;Keywords: chloroplast

Query Match 50.6%; Score 43; DB 2; Length 2054;
Best Local Similarity 46.7%; Pred. No. 77;
Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

```
Qy       2 NLVPMNPLVMNEFMN 16
           | :|:|||::: | |
Db     717 NKIPINPLIIDFFDN 731
```

RESULT 3
F75457
conserved hypothetical protein - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C;Accession: F75457
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma .; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896
A;Accession: F75457
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-346 <WHI>
A;Cross-references: GB:AE001946; GB:AE000513; NID:g6458655; PIDN:AAF10513.1; PID:g645866
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR0936
A;Map position: 1
C;Superfamily: conserved hypothetical protein HI0365

Query Match 49.4%; Score 42; DB 2; Length 346;
Best Local Similarity 43.8%; Pred. No. 14;
Matches 7; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

```
Qy       1 MNLVPMNPLVMNEFMN 16
             :||:|||| ::|::
Db     279 VNLIPMNPWDGSDFVS 294
```

RESULT 4
VPXRHK
outer layer protein VP3 - human rotavirus A (strain KU)
N;Alternate names: glycoprotein VP3; hemagglutinin; outer capsid protein VP3
C;Species: human rotavirus A
C;Date: 30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change 16-Jul-1999
C;Accession: A28844
R;Taniguchi, K.; Maloy, W.L.; Nishikawa, K.; Green, K.Y.; Hoshino, Y.; Urasawa, S.; Kapi
J. Virol. 62, 2421-2426, 1988
A;Title: Identification of cross-reactive and serotype 2-specific neutralization epitope
A;Reference number: A28844; MUID:88230603
A;Accession: A28844
A;Molecule type: mRNA
A;Residues: 1-775 <TAN>
A;Cross-references: GB:M21014; NID:g333852; PIDN:AAA47334.1; PID:g333853
C;Genetics:
A;Map position: segment 4
C;Superfamily: rotavirus outer layer protein VP3
C;Keywords: glycoprotein; hemagglutinin; outer capsid protein
F;32,56,97,132,324,583,589,592,599/Binding site: carbohydrate (Asn) (covalent) #status p

Query Match 49.4%; Score 42; DB 1; Length 775;
Best Local Similarity 43.8%; Pred. No. 36;
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

```
Qy       1 MNLVPMNPLVMNEFMN 16
             :||:  || |:  |:|
Db     741 LNLIKSNPNVLRNFIN 756
```

RESULT 5
VPXRWL
outer layer protein VP3 - human rotavirus A (strain L26)
N;Alternate names: VP4 protein
N;Contains: outer capsid protein VP5; outer capsid protein VP8
C;Species: human rotavirus A
C;Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 16-Jul-1999
C;Accession: A36410
R;Taniguchi, K.; Urasawa, T.; Kobayashi, N.; Gorziglia, M.; Urasawa, S.
J. Virol. 64, 5640-5644, 1990
A;Title: Nucleotide sequence of VP4 and VP7 genes of human rotaviruses with subgroup
A;Reference number: A36410; MUID:91012813
A;Accession: A36410
A;Molecule type: genomic RNA
A;Residues: 1-775 <TAN>
A;Cross-references: EMBL:M58292; NID:g333854; PIDN:AAA47335.1; PID:g333855
C;Genetics:
A;Map position: segment 4
C;Superfamily: rotavirus outer layer protein VP3
C;Keywords: capsid protein; glycoprotein
F;1-240/Product: outer capsid protein VP8 #status predicted <VP8>
F;241-246/Region: cleavage processing #status predicted
F;247-775/Product: outer capsid protein VP5 #status predicted <VP5>
F;32,56,97,132,150,195,324,583,589,599/Binding site: carbohydrate (Asn) (covalent) #s

Query Match 49.4%; Score 42; DB 1; Length 775;
Best Local Similarity 43.8%; Pred. No. 36;
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

```
Qy       1 MNLVPMNPLVMNEFMN 16
             :||:  || |:  |:|
Db     741 LNLIKSNPNVLRNFIN 756
```

RESULT 6
VPXRWM
outer layer protein VP3 - human rotavirus A (strain L27)
N;Alternate names: VP4 protein
N;Contains: outer capsid protein VP5; outer capsid protein VP8
C;Species: human rotavirus A
C;Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 01-Mar-1996
C;Accession: C36410
R;Taniguchi, K.; Urasawa, T.; Kobayashi, N.; Gorziglia, M.; Urasawa, S.
J. Virol. 64, 5640-5644, 1990
A;Title: Nucleotide sequence of VP4 and VP7 genes of human rotaviruses with subgroup
A;Reference number: A36410; MUID:91012813
A;Accession: C36410
A;Molecule type: genomic RNA
A;Residues: 1-775 <TAN>
A;Cross-references: EMBL:M58292
C;Genetics:
A;Map position: segment 4
C;Superfamily: rotavirus outer layer protein VP3
C;Keywords: capsid protein; glycoprotein
F;1-240/Product: outer capsid protein VP8 #status predicted <VP8>
F;241-246/Region: cleavage processing #status predicted
F;247-775/Product: outer capsid protein VP5 #status predicted <VP5>
F;32,56,97,132,150,195,324,583,589,599/Binding site: carbohydrate (Asn) (covalent) #s

Query Match 49.4%; Score 42; DB 1; Length 775;
Best Local Similarity 43.8%; Pred. No. 36;
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

```
Qy       1 MNLVPMNPLVMNEFMN 16
             :||:  || |:  |:|
Db     741 LNLIKSNPNVLRNFIN 756
```

RESULT 7

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: January 6, 2002, 11:02:28 ; Search time 121.65 Seconds
(without alignments)
94.948 Million cell updates/sec

Title: US-09-897-042-17
Perfect score: 51
Sequence: 1 atgaacttggtaccgatgaa.....tgaacgaattcatgaactaa 51

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	21.4	42.0	1912	1	US-08-270-013B-1	Sequence 1, Appli
	2	21.4	42.0	1912	1	US-08-838-418-1	Sequence 1, Appli
c	3	21.2	41.6	805	3	US-08-961-083-139	Sequence 139, App
	4	21.2	41.6	1945	1	US-08-724-194-1	Sequence 1, Appli
	5	21.2	41.6	2678	1	US-08-724-194-2	Sequence 2, Appli
	6	20.8	40.8	408	4	US-09-328-111-169	Sequence 169, App
c	7	20.8	40.8	3063	2	US-08-184-009-169	Sequence 169, App
c	8	20.8	40.8	3063	2	US-08-458-356-169	Sequence 169, App
c	9	20.8	40.8	3063	4	US-08-460-736-169	Sequence 169, App
c	10	20.8	40.8	3706	2	US-08-566-398-59	Sequence 59, Appl
c	11	20.8	40.8	3706	2	US-08-658-665-63	Sequence 63, Appl
c	12	20.8	40.8	3706	4	US-08-796-101-27	Sequence 27, Appl
c	13	20.8	40.8	3706	4	US-09-085-273-63	Sequence 63, Appl
	14	20.6	40.4	176	2	US-08-687-080-82	Sequence 82, Appl
	15	20.6	40.4	486	3	US-09-217-609A-28	Sequence 28, Appl
	16	20.6	40.4	486	4	US-08-873-235B-28	Sequence 28, Appl
c	17	20.6	40.4	802	4	US-08-998-416-376	Sequence 376, App
	18	20.6	40.4	1243	2	US-08-687-080-53	Sequence 53, Appl
	19	20.6	40.4	5893	1	US-08-592-126-54	Sequence 54, Appl
	20	20.6	40.4	5893	2	US-08-687-080-44	Sequence 44, Appl
c	21	20.4	40.0	420	4	US-09-214-095D-115	Sequence 115, App
	22	20.4	40.0	423	1	US-08-470-179-120	Sequence 120, App
c	23	20.4	40.0	1854	1	US-08-249-420-1	Sequence 1, Appli
c	24	20.4	40.0	1854	2	US-08-737-663-1	Sequence 1, Appli
	25	20.2	39.6	638	1	US-08-469-667-1	Sequence 1, Appli
	26	20.2	39.6	638	5	PCT-US95-07289-1	Sequence 1, Appli
c	27	20.2	39.6	56516	2	US-08-996-306-1	Sequence 1, Appli
c	28	20.2	39.6	56516	4	US-09-338-907-1	Sequence 1, Appli
c	29	20.2	39.6	56520	4	US-09-338-907-179	Sequence 179, App
	30	19.8	38.8	423	1	US-08-470-179-39	Sequence 39, Appl
	31	19.8	38.8	423	1	US-08-470-179-45	Sequence 45, Appl
	32	19.8	38.8	423	1	US-08-470-179-53	Sequence 53, Appl
	33	19.8	38.8	899	1	US-07-820-154A-3	Sequence 3, Appli
	34	19.8	38.8	899	2	US-08-097-554A-3	Sequence 3, Appli
	35	19.8	38.8	899	3	US-08-480-640A-3	Sequence 3, Appli
	36	19.8	38.8	899	3	US-08-295-802-3	Sequence 3, Appli
	37	19.8	38.8	899	4	US-08-488-237A-3	Sequence 3, Appli
	38	19.8	38.8	899	5	PCT-US93-00324-3	Sequence 3, Appli
	39	19.8	38.8	3628	3	US-08-480-640A-113	Sequence 113, App
	40	19.8	38.8	3628	3	US-08-295-802-113	Sequence 113, App
	41	19.8	38.8	3628	4	US-08-488-237A-113	Sequence 113, App
	42	19.8	38.8	3942	3	US-08-480-640A-189	Sequence 189, App
	43	19.8	38.8	3942	4	US-08-686-968C-189	Sequence 189, App
	44	19.8	38.8	3942	4	US-08-488-237A-189	Sequence 189, App
	45	19.8	38.8	5785	3	US-08-480-640A-221	Sequence 221, App

ALIGNMENTS

```
RESULT   1
US-08-270-013B-1
; Sequence 1, Application US/08270013B
; Patent No. 5686294
;  GENERAL INFORMATION:
;    APPLICANT:  Sogabe et al.
;    TITLE OF INVENTION:  PROTEIN HAVING HEAT-RESISTANT MALATE
;    TITLE OF INVENTION:  DEHYDROGENASE ACTIVITY
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Leydig, Voit & Mayer, Ltd.
;      STREET:  Two Prudential Plaza, Suite 4900
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  61601-6780
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/270,013B
;      FILING DATE:  01-JUL-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 164701/1993
;      FILING DATE:  02-JUL-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Green, Robert F.
;      REGISTRATION NUMBER:  27555
;      REFERENCE/DOCKET NUMBER:  62321
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (312) 616-5600
;      TELEFAX:  (312) 616-5700
;      TELEX:  (25)3533
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1912 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Bacillus stearothermophilus
;      STRAIN:  ATCC12016
US-08-270-013B-1
```

```
  Query Match             42.0%;  Score 21.4;  DB 1;  Length 1912;
  Best Local Similarity   66.0%;  Pred. No. 11;
  Matches   31;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy       5 acttggtaccgatgaacccactagtaatgaacgaattcatgaactaa 51
           | |||||  ||| ||||| |  | | |||||  || | | |||  ||
Db     966 ATTTGGTGACGACGAACCAAAAAATTATGAAGCAAGTGACGAAGGAA 1012


RESULT   2
US-08-838-418-1
; Sequence 1, Application US/08838418
; Patent No. 5744342
;  GENERAL INFORMATION:
;    APPLICANT:  Sogabe et al.
;    TITLE OF INVENTION:  PROTEIN HAVING HEAT-RESISTANT MALATE
;    TITLE OF INVENTION:  DEHYDROGENASE ACTIVITY
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Leydig, Voit & Mayer, Ltd.
;      STREET:  Two Prudential Plaza, Suite 4900
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60601-6780
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/838,418
;      FILING DATE:  17-MAR-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/270,013
;      FILING DATE:  01-JUL-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 164701/1993
;      FILING DATE:  02-JUL-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hoover, Allen E.
;      REGISTRATION NUMBER:  37354
;      REFERENCE/DOCKET NUMBER:  78339
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (312) 616-5600
;      TELEFAX:  (312) 616-5700
;      TELEX:  (25)3533
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1912 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Bacillus stearothermophilus
;      STRAIN:  ATCC12016
US-08-838-418-1


  Query Match             42.0%;  Score 21.4;  DB 1;  Length 1912;
  Best Local Similarity   66.0%;  Pred. No. 11;
  Matches   31;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy       5 acttggtaccgatgaacccactagtaatgaacgaattcatgaactaa 51
           | |||||  ||| ||||| |  | | |||||  || | | |||  ||
Db     966 ATTTGGTGACGACGAACCAAAAAATTATGAAGCAAGTGACGAAGGAA 1012


RESULT   3
US-08-961-083-139/c
; Sequence 139, Application US/08961083
; Patent No. 6159469
;  GENERAL INFORMATION:
;    APPLICANT:  Choi et. al.
;    TITLE OF INVENTION:  Streptococcus pneumoniae Antigens and Vaccines
;    NUMBER OF SEQUENCES:  452
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/961,083
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brookes, A. Anders
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PB340P2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  139:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  805 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-961-083-139


  Query Match             41.6%;  Score 21.2;  DB 3;  Length 805;
  Best Local Similarity   64.0%;  Pred. No. 11;
  Matches   32;  Conservative    0;  Mismatches   18;  Indels    0;  Gaps    0;

Qy       1 atgaacttggtaccgatgaacccactagtaatgaacgaattcatgaacta 50
           |||||  | || | ||||| ||  |   |  ||||| |  || ||||| |
Db     638 ATGAATATTGTTCTGATGATCCACCATTTGTTGAACTATCTCCTGAACCA 589


RESULT   4
US-08-724-194-1/c
; Sequence 1, Application US/08724194
; Patent No. 5824875
;  GENERAL INFORMATION:
;    APPLICANT:  RANU, RAJINDER S.
;    TITLE OF INVENTION:  ONE-AMINOCYCLOPROPANE-1-CARBOXYLATE
;    TITLE OF INVENTION:  SYNTHASE GENES FROM PELARGONIUM TO CONTROL ETHYLENE LEVELS
;    TITLE OF INVENTION:  IN GERANIUMS
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SANTANGELO LAW OFFICES PC
;      STREET:  315 WEST OAK STREET, STE 701
;      CITY:  FORT COLLINS
;      STATE:  CO
;      COUNTRY:  USA
;      ZIP:  80521
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: January 6, 2002, 11:06:56 ; Search time 256.95 Seconds
(without alignments)
170.164 Million cell updates/sec

Title: US-09-897-042-17
Perfect score: 51
Sequence: 1 atgaacttggtaccgatgaa.....tgaacgaattcatgaactaa 51

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_1101:*
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	24.4	47.8	3246	22	AAH53972	S. epidermidis gen
	2	24	47.1	630	20	AAX61539	B. burgdorferi ant
	3	24	47.1	116277	20	AAX20249	Borrelia burgdorfe
c	4	23.2	45.5	2244	12	AAQ11658	3' Terminal of TSW
c	5	23.2	45.5	6706	22	AAC89648	Tomato spotted wil
c	6	23.2	45.5	6708	20	AAX86463	Partial sequence o
c	7	23.2	45.5	8897	20	AAX86464	Nucleotide sequenc
c	8	23.2	45.5	8897	22	AAC89649	Tomato spotted wil
	9	22.8	44.7	4191	19	AAV23692	L.lactis MG1363 pf
	10	22.6	44.3	7207	21	AAF21316	Human low adenosin
	11	22.6	44.3	7207	21	AAA35194	Human adenosine re
	12	22.6	44.3	15630	21	AAF21317	Human low adenosin
	13	22.6	44.3	15630	21	AAA35195	Human adenosine re
	14	22.2	43.5	4783	21	AAC76742	Human ORFX ORF2297
	15	22	43.1	1830121	17	AAT42063	Haemophilus influe
c	16	21.8	42.7	371	22	AAF67075	Novel human polynu
	17	21.8	42.7	558	22	AAH10735	Human cDNA clone (
c	18	21.8	42.7	1599	21	AAA97797	Tetragenococcus ha
c	19	21.8	42.7	2683	22	AAH17616	Human cDNA sequenc
c	20	21.8	42.7	4115	21	AAA97430	Gluconobacter oxyd
	21	21.6	42.4	245	20	AAZ13521	Human gene express
	22	21.6	42.4	300	20	AAZ13737	Human gene express
	23	21.6	42.4	300	20	AAX98348	Human cancer cell
c	24	21.6	42.4	551	22	AAH11430	Human cDNA clone (
	25	21.6	42.4	753	20	AAZ15437	Human gene express
	26	21.6	42.4	1761	21	AAZ45339	DNA encoding a Bac
	27	21.6	42.4	2522	22	AAH16195	Human cDNA sequenc
c	28	21.4	42.0	171	22	AAI61744	Soybean 515002 reg
c	29	21.4	42.0	171	22	AAI61745	Soybean 515002 reg
	30	21.4	42.0	510	21	AAC50790	Arabidopsis thalia
	31	21.4	42.0	1132	21	AAC38850	Arabidopsis thalia
	32	21.4	42.0	1912	17	AAT17715	Heat resistant mal
c	33	21.4	42.0	2607	19	AAV19464	C. utilis INV1 gen
c	34	21.4	42.0	127197	22	AAI61370	Soybean 515002 reg
	35	21.4	42.0	335913	22	AAI61371	Soybean 240017 reg
	36	21.4	42.0	335913	22	AAI61372	Soybean 240017 reg
	37	21.2	41.6	322	20	AAV87277	EST clone BO535.
c	38	21.2	41.6	805	19	AAV27393	Streptococcus pneu
c	39	21.2	41.6	1617	22	AAH53870	S. epidermidis ope
c	40	21.2	41.6	1934	18	AAT66246	ACC synthase GAC-1
c	41	21.2	41.6	1945	19	AAV30324	Pelargonium 1-amin
c	42	21.2	41.6	2678	19	AAV30325	Pelargonium 1-amin
	43	21.2	41.6	3073	22	AAH54242	S. epidermidis gen
	44	21.2	41.6	3723	22	AAH54274	S. epidermidis gen
	45	21.2	41.6	4323	22	AAI58257	Human polynucleoti

ALIGNMENTS

RESULT 1
AAH53972
ID AAH53972 standard; DNA; 3246 BP.
XX
AC AAH53972;
XX
DT 03-SEP-2001 (first entry)
XX
DE S. epidermidis genomic polynucleotide sequence SEQ ID NO:3336.
XX
KW Staphylococcus epidermidis SR1 strain; infection; diagnosis;
KW vaccination; endocarditis; ds.
XX
OS Staphylococcus epidermidis.
XX
PN WO200134809-A2.
XX
PD 17-MAY-2001.
XX
PF 09-NOV-2000; 2000WO-US30782.
XX
PR 09-NOV-1999; 99US-0164258.
XX
PA (GLAX) GLAXO GROUP LTD.
XX
PI Kimmerly WJ;
XX
DR WPI; 2001-316495/33.
XX
PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
PT useful for vaccinating against infections, e.g. endocarditis -
XX
PS Claim 8; Page 881-882; 2188pp; English.
XX

CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
CC (I) and (II) can have antibacterial activity and therefore can be used
CC in vaccination. The nucleic acids (I) may be used to produce the
CC S. epidermidis polypeptides (II) via the production of vectors
CC containing them which are used to produce hosts cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to
CC AAH55098 represent oligonucleotide sequences and primers which are used
CC in the exemplification of the present invention.
CC N.B. The present invention specifically claims all the polynucleotide
CC sequences given in the sequence listing of the present specification,
CC however the sequence listing only goes up to SEQ ID NO:4454 so even
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC no sequences are present for SEQ ID NO:4455 to 4464.
XX
SQ Sequence 3246 BP; 1091 A; 533 C; 536 G; 1086 T; 0 other;

Query Match 47.8%; Score 24.4; DB 22; Length 3246;
Best Local Similarity 68.0%; Pred. No. 4.3;
Matches 34; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy 1 atgaacttggtaccgatgaacccactagtaatgaacgaattcatgaacta 50
||| |||||| || |||| | || |||||||||||| ||| ||
Db 871 atggttttggtaaagaagaacaacatttaaaagaacgaattcaagaatta 920

RESULT 2
AAX61539
ID AAX61539 standard; DNA; 630 BP.
XX
AC AAX61539;
XX
DT 19-JUL-1999 (first entry)
XX
DE B. burgdorferi antigenic protein coding sequence, f516.nt.
XX
KW Antigenic protein; vaccine; Lyme disease; infection; detection; ss.
XX
OS Borrelia burgdorferi.
XX
PN WO9859071-A1.
XX
PD 30-DEC-1998.
XX
PF 18-JUN-1998; 98WO-US12718.
XX
PR 03-SEP-1997; 97US-0057483.
PR 20-JUN-1997; 97US-0050359.
PR 22-JUL-1997; 97US-0053344.
PR 22-JUL-1997; 97US-0053377.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (MEDI-) MEDIMMUNE INC.
XX
PI Choi GH, Erwin AL, Hanson MS, Lathigra R;
XX
DR WPI; 1999-189980/16.
DR P-PSDB; AAY19842.
XX
PT New isolated Borrelia burgdorferi nucleic acids - used to develop
PT products for the diagnosis, prevention and treatment of diseases
PT caused by Borrelia, particularly Lyme disease
XX
PS Claim 1; Page 88; 275pp; English.
XX
CC This sequence encodes a Borrelia burgdorferi (Bb) protein of the
CC invention, which is suitable for use in a vaccine. The Bb polypeptides
CC can be used in vaccines for eliciting protective antibodies to members of
CC the Borrelia genus, particularly for the use against Lyme disease in
CC humans and animals. They can be used for preventing or attenuating an
CC infection caused by a member of the Borrelia genus. The products can also
CC be used for detection of members of the Borrelia genus.
XX
SQ Sequence 630 BP; 233 A; 92 C; 81 G; 224 T; 0 other;

Query Match 47.1%; Score 24; DB 20; Length 630;
Best Local Similarity 68.8%; Pred. No. 4.6;
Matches 33; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 1 atgaacttggtaccgatgaacccactagtaatgaacgaattcatgaac 48
||| ||||| | || ||| || || ||||||| || | || |||
Db 59 atgcacttgcaaatgaagaaggcaatactaatgaaaaaaatgatcaac 106

RESULT 3
AAX20249
ID AAX20249 standard; DNA; 116277 BP.
XX
AC AAX20249;
XX
DT 04-MAY-1999 (first entry)
XX
DE Borrelia burgdorferi polynucleotide sequence #2.
XX
KW Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;
KW epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
KW infection; diagnosis; characterisation; detection; ds.
XX
OS Borrelia burgdorferi.
XX
PN WO9858943-A1.
XX
PD 30-DEC-1998.
XX
PF 18-JUN-1998; 98WO-US12764.
XX
PR 03-SEP-1997; 97US-0057483.
PR 20-JUN-1997; 97US-0050359.
PR 22-JUL-1997; 97US-0053344.
PR 22-JUL-1997; 97US-0053377.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (MEDI-) MEDIMMUNE INC.
XX
PI Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO;
PI White OR;
XX
DR WPI; 1999-081217/07.
XX
PT New isolated Borrelia burgdorferi nucleic acids - used to develop
PT products for the detection, diagnosis, characterisation, prevention
PT and therapy of infections, particularly Lyme disease
XX
PS Claim 1; Page 672-737; 1128pp; English.
XX
CC AAX20248 to AAX20402 represent polynucleotide sequences isolated from
CC Borrelia burgdorferi (Bb). Products derived from Bb can be used for
CC the detection, diagnosis, characterisation, prevention and therapy of
CC Bb infections, e.g. Lyme disease. They can also be used for the
CC production of biosynthetic products, e.g. enzymes. Borrelia belongs
CC to a family of motile, spiral-shaped bacteria called Spirochetes.
CC Spirochetes are pathogenic in humans and Borrelia causes epidemic and
CC endemic relapsing fever, and Lyme borreliosis, more commonly known as
CC Lyme disease.
XX
SQ Sequence 116277 BP; 42656 A; 19868 C; 14490 G; 39250 T; 13 other;

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: January 6, 2002, 10:33:30 ; Search time 2642.04 Seconds
(without alignments)
207.429 Million cell updates/sec

Title: US-09-897-042-17
Perfect score: 51
Sequence: 1 atgaacttggtaccgatgaa.....tgaacgaattcatgaactaa 51

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_htc:*
10: gb_est1:*
11: gb_est2:*
12: gb_htc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	25.8	50.6	511	11	BG041695	BG041695 sv37g01.y
c	2	25	49.0	607	11	C84717	C84717 C84717 Dict
	3	24.8	48.6	465	10	AW641204	AW641204 cm04e04.w
c	4	24.6	48.2	883	13	CNS07D91	AL440171 T3 end of
	5	24.4	47.8	348	10	AV548944	AV548944 AV548944
c	6	24.4	47.8	506	10	AV543742	AV543742 AV543742
c	7	24.4	47.8	529	10	AV544086	AV544086 AV544086
c	8	24.4	47.8	538	13	AQ968417	AQ968417 LERJC35TR
c	9	24.4	47.8	655	10	AV545019	AV545019 AV545019
	10	24.4	47.8	662	13	AQ968416	AQ968416 LERJC35TF
c	11	24.2	47.5	682	13	AQ795238	AQ795238 nbxb0055N
c	12	24.2	47.5	761	13	AQ326076	AQ326076 nbxb0024J
c	13	24.2	47.5	776	13	AQ576890	AQ576890 nbxb0089P
	14	24	47.1	843	13	CNS03BRH	AL236870 Tetraodon
	15	23.8	46.7	392	10	AW371950	AW371950 RC4-BT031
c	16	23.8	46.7	419	13	AQ563015	AQ563015 HS_5301_A
	17	23.8	46.7	1070	13	CNS06PEX	AL409279 T3 end of
	18	23.6	46.3	365	11	BF415704	BF415704 UI-R-CA1-
	19	23.6	46.3	528	11	BF416556	BF416556 UI-R-CA1-
c	20	23.6	46.3	535	13	AZ467538	AZ467538 1M0279B04
	21	23.6	46.3	570	13	AZ371615	AZ371615 1M0123D12
	22	23.6	46.3	575	13	AZ371859	AZ371859 1M0123P22
c	23	23.6	46.3	793	10	BE130386	BE130386 L48-560T3
c	24	23.6	46.3	964	11	BG387522	BG387522 602412333
	25	23.6	46.3	1187	11	BF796700	BF796700 602258321
c	26	23.4	45.9	363	13	AQ033770	AQ033770 HS_2237_A
c	27	23.4	45.9	471	11	N91806	N91806 zb45a08.s1
	28	23.4	45.9	496	11	BG318404	BG318404 NXPV_013_
	29	23.4	45.9	537	11	BI202938	BI202938 NXPV_092_
	30	23.4	45.9	557	11	BG318869	BG318869 NXPV_020_
	31	23.4	45.9	595	10	AA495617	AA495617 c428 Zhou
	32	23.4	45.9	967	13	CNS06PL6	AL409504 T3 end of
	33	23.2	45.5	272	11	BG124644	BG124644 EST470290
c	34	23.2	45.5	404	11	BG629894	BG629894 cC-esf1cL
c	35	23.2	45.5	522	11	BF112629	BF112629 EST440219
c	36	23	45.1	265	10	BB592896	BB592896 BB592896
	37	23	45.1	560	11	BI424347	BI424347 saf31g03.
	38	23	45.1	768	13	AQ420780	AQ420780 RPCI-11-2
	39	23	45.1	844	13	AZ534572	AZ534572 ENTCS77TR
c	40	22.8	44.7	255	10	AI071750	AI071750 UI-R-C2-n
c	41	22.8	44.7	307	11	N59695	N59695 yv55g05.r1
c	42	22.8	44.7	459	10	AV542740	AV542740 AV542740
	43	22.8	44.7	467	10	BE587941	BE587941 WHE0665_F
c	44	22.8	44.7	562	11	BG135904	BG135904 EST476346
	45	22.8	44.7	567	13	AZ711369	AZ711369 RPCI-24-1

ALIGNMENTS

RESULT 1
BG041695
LOCUS BG041695 511 bp mRNA EST 31-JUL-2001
DEFINITION sv37g01.y1 Gm-c1057 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1057-1849 5' similar to TR:Q9SH32 Q9SH32 F2K11.16. ;, mRNA sequence.
ACCESSION BG041695
VERSION BG041695.1 GI:12487985
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
REFERENCE 1 (bases 1 to 511)
AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann ,R., Waterston,R. and Wilson,R.
TITLE Public Soybean EST Project
JOURNAL Unpublished (1999)
COMMENT Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or

```
          info@genomesystems.com web site: www.genomesystems.com
          Insert Length: 837   Std Error: 0.00
          High quality sequence stop: 416.
FEATURES             Location/Qualifiers
     source          1. .511
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1057-1849"
                     /clone_lib="Gm-c1057"
                     /tissue_type="Degenerating cotyledons, 2 week old
                     seedling"
                     /lab_host="DH10B"
                     /note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
                     XhoI; The cDNA library was constructed from mRNA isolated
                     from degenerating cotyledons of 2 week old seedlings from
                     PI468916. Complementary DNA was  synthesized from mRNA
                     using a primer consisting of a poly(dT) sequence with  a
                     XhoI restriction site. EcoRI adapters were ligated to the
                     blunt-ended cDNA  fragments followed by XhoI digestion.
                     The cDNA fragments were directionally  cloned into the
                     EcoRI-XhoI restriction site of the pBluescript vector. The
                     ligated cDNA fragments were transformed into DH10B host
                     cells (GibcoBRL).  This library was constructed in the
                     laboratory of Dr. Randy Shoemaker."
BASE COUNT      141 a     96 c    122 g    150 t      2 others
ORIGIN


  Query Match             50.6%;  Score 25.8;  DB 11;  Length 511;
  Best Local Similarity   71.7%;  Pred. No. 42;
  Matches   33;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy      1 atgaacttggtaccgatgaacccactagtaatgaacgaattcatga 46
          || || |||||   || ||||||   || ||||| | |||||| ||
Db    231 ATCAATTTGGTGAGGAGGAACCCGGAAGCAATGATCAAATTCANGA 276


RESULT   2
C84717/c
LOCUS       C84717        607 bp    mRNA            EST       12-JUL-1999
DEFINITION  C84717 Dictyostelium discoideum SS (H.Urushihara) Dictyostelium
            discoideum cDNA clone SSE735, mRNA sequence.
ACCESSION   C84717
VERSION     C84717.1  GI:2864864
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 607)
  AUTHORS   Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
            Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
            Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
  TITLE     The Dictyostelium developmental cDNA project: generation and
            analysis of expressed sequence tags from the first-finger stage of
            development
  JOURNAL   DNA Res. 5 (6), 335-340 (1998)
  MEDLINE   99156227
COMMENT     Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan
            Email: d402hu@sakura.cc.tsukuba.ac.jp
            PROJECT = 'Dictyostelium discoideum cDNA project in Japan'
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .607
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="SSE735"
                     /clone_lib="Dictyostelium discoideum SS (H.Urushihara)"
                     /dev_stage="slug"
BASE COUNT      215 a    103 c     67 g    222 t
ORIGIN


  Query Match             49.0%;  Score 25;  DB 11;  Length 607;
  Best Local Similarity   69.4%;  Pred. No. 80;
  Matches   34;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy      3 gaacttggtaccgatgaacccactagtaatgaacgaattcatgaactaa 51
          ||||||| |   | ||||||  ||| |  ||||| | ||| |||| | |
Db    411 GAACTTGATGTTGTTGAACCACCTATTGTTGAACCACTTCCTGAAGTTA 363


RESULT   3
AW641204
LOCUS       AW641204      465 bp    mRNA            EST       26-APR-2001
DEFINITION  cm04e04.w1 Blackshear/Soares normalized Xenopus egg library Xenopus
            laevis cDNA clone PBX0103E04 5', mRNA sequence.
ACCESSION   AW641204
VERSION     AW641204.1  GI:7398457
KEYWORDS    EST.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 465)
  AUTHORS   Blackshear,P.J., Lai,W.S., Thorn,J.M., Kennington,E.A., Staffa,N.G.
            Jr., Moore,D.T., Bouffard,G.G., Beckstrom-Sternberg,S.M., Touchman
            ,J.W., Bonaldo,M.F. and Soares,M.B.
  TITLE     The NIEHS Xenopus maternal EST project: interim analysis of the
            first 13,879 ESTs from unfertilized eggs
  JOURNAL   Gene 267 (1), 71-87 (2001)
  MEDLINE   21211403
COMMENT     Contact: Perry J. Blackshear
            Office of Clinical Research and Laboratory of Signal Transduction
            National Institute of Environmental Health Sciences
            A2-05 NIEHS, 101 Alexander Drive, Research Triangle Park, NC 27709,
            USA
            Tel: 919 541-4899
            Fax: 919 541-4571
            Email: black009@niehs.nih.gov
            Clone is available through Research Genetics, Inc., 2130 Memorial
            Parkway, Huntsville, AL 35901
             phone 800-533-4363 ext.cdna, fax 256-536-9016 att:cdna, email
            cdna@resgen.com
             DNA Sequencing and analyses performed by National Institutes of
            Health Intramural Sequencing Center (NISC).
            PCR PRimers
            FORWARD: TGTAAAACGACGGCCAGT
            BACKWARD: CAGGAAACAGCTATGACC
            Plate: 0103  row: E  column: 04
            Seq primer: T7 primer.
FEATURES             Location/Qualifiers
     source          1. .465
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /clone="PBX0103E04"
                     /clone_lib="Blackshear/Soares normalized Xenopus egg
                     library"
                     /sex="female"
                     /tissue_type="unfertilized egg"
                     /cell_type="unfertilized egg"
                     /dev_stage="unfertilized egg"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3-Pac; Site_1: EcoRI; Site_2: NotI;
                     PolyA-selected mRNA was prepared from unfertilized Xenopus
                     laevis eggs.  The library was constructed in the vector
                     pT7T3-Pac as described in Bonaldo, M.F., Lennon, G. and
                     Soares, M.B. 'Normalization and subtraction:  two
                     approaches to facilitate gene discovery',  Genome Research
                     6:791-806, 1996.  The first strand synthesis used a
                     NotI-dT18 primer; double stranded cDNAs were ligated to
                     EcoRI adapters, digested with NotI, and directionally
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: January 6, 2002, 11:00:38 ; Search time 1586.23 Seconds
(without alignments)
530.412 Million cell updates/sec

Title: US-09-897-042-17
Perfect score: 51
Sequence: 1 atgaacttggtaccgatgaa.....tgaacgaattcatgaactaa 51

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vi:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rod:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rod:*
36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

%

Result No.		Score	Query Match	Length	DB	ID	Description
c	1	26.4	51.8	8601	3	AF274754	AF274754 Brugia ma
	2	24.8	48.6	168497	2	AC092410	AC092410 Bos tauru
	3	24.6	48.2	4536	14	FIPPP	D32044 Feline infe
c	4	24.6	48.2	300950	1	AP001516	AP001516 Bacillus
	5	24.4	47.8	3246	1	AF269294	AF269294 Staphyloc
	6	24.4	47.8	3246	6	AX144614	AX144614 Sequence
c	7	24.4	47.8	57198	2	AC020115	AC020115 Drosophil
c	8	24.4	47.8	118507	8	ATF7J8	AL137189 Arabidops
	9	24.4	47.8	169090	3	AC007300	AC007300 Drosophil
	10	24.4	47.8	260791	3	AE003629	AE003629 Drosophil
	11	24	47.1	18604	1	AE001146	AE001146 Borrelia
	12	24	47.1	35050	2	AC006099	AC006099 Homo sapi
	13	24	47.1	67249	2	AC018611	AC018611 Homo sapi
c	14	24	47.1	67249	2	AC018611	AC018611 Homo sapi
	15	24	47.1	163729	2	AC013252	AC013252 Homo sapi
c	16	24	47.1	167501	2	AL391421	AL391421 Homo sapi
c	17	24	47.1	171505	9	CNS01DTH	AL132719 Human chr
	18	24	47.1	180506	2	AC069306	AC069306 Homo sapi
	19	24	47.1	219476	9	AC068945	AC068945 Homo sapi
	20	23.6	46.3	6380	8	PETNIAA	L11563 Petunia hyb
	21	23.6	46.3	6380	8	PETNITRED	L13691 Petunia hyb
c	22	23.6	46.3	134534	2	AL512352	AL512352 Homo sapi
c	23	23.6	46.3	152027	2	AC027386	AC027386 Homo sapi
	24	23.6	46.3	157657	9	AC022841	AC022841 Homo sapi
c	25	23.6	46.3	174494	2	AC068744	AC068744 Homo sapi
	26	23.6	46.3	190056	2	AP002430	AP002430 Homo sapi
c	27	23.6	46.3	190642	2	AC062013	AC062013 Homo sapi
c	28	23.4	45.9	3413	5	CHA249579	AJ249579 Chionodra
	29	23.4	45.9	166036	2	AC041013	AC041013 Homo sapi
	30	23.4	45.9	179702	2	AC090058	AC090058 Homo sapi
	31	23.4	45.9	205363	2	AC078998	AC078998 Mus muscu
	32	23.2	45.5	2239	8	SLU60267	U60267 Solanum lyc
	33	23.2	45.5	8897	14	TSWLRPOLM	D10066 Tomato spot
	34	23.2	45.5	182056	9	AC073533	AC073533 Homo sapi
c	35	23	45.1	23830	8	SPAC27F1	Z69368 S.pombe chr
c	36	23	45.1	40029	3	CEF25C8	Z81512 Caenorhabdi
c	37	23	45.1	43160	2	CEF14H10	Z92783 Caenorhabdi
	38	23	45.1	62844	2	AC023982	AC023982 Homo sapi
c	39	23	45.1	127590	9	AC002554	AC002554 Human Chr
c	40	23	45.1	152567	33	AC011834	AC011834 Homo sapi
	41	23	45.1	161737	9	AC090512	AC090512 Homo sapi
	42	23	45.1	174194	2	AL512329	AL512329 Homo sapi
	43	22.8	44.7	2600	4	EKU97342	U97342 Echymipera
	44	22.8	44.7	4191	1	LLPFLMG13	AJ000325 Lactococc
	45	22.8	44.7	4481	3	DGU94491	U94491 Dictyosteli

ALIGNMENTS

RESULT 1
AF274754/c
LOCUS AF274754 8601 bp DNA INV 23-JAN-2001
DEFINITION Brugia malayi chitin synthase (chs1) gene, complete cds.
ACCESSION AF274754
VERSION AF274754.1 GI:12382255
KEYWORDS .
SOURCE Brugia malayi.
ORGANISM Brugia malayi
Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea; Onchocercidae; Brugia.
REFERENCE 1 (bases 1 to 8601)
AUTHORS Harris,M.T., Lai,K., Arnold,K., Martinez,H.F., Specht,C.A. and Fuhrman,J.A.
TITLE Chitin synthase in the filarial parasite, Brugia malayi
JOURNAL Mol. Biochem. Parasitol. 111 (2), 351-362 (2000)
MEDLINE 21036600
REFERENCE 2 (bases 1 to 8601)
AUTHORS Fuhrman,J.A., Arnold,K., Lai,K. and Martinez,H.F.
TITLE Direct Submission

```
  JOURNAL   Submitted (02-JUN-2000) Biology, Tufts University, Dana Labs,
            Medford, MA 02155, USA
FEATURES             Location/Qualifiers
     source          1. .8601
                     /organism="Brugia malayi"
                     /specific_host="Meriones unguiculatus"
                     /db_xref="taxon:6279"
                     /note="subperiodic Brugia malayi from i.p. infected
                     Meriones unguiculatus; obtained from the Filariasis
                     Repository Research Service, Univ. of Georgia"
     misc_feature    794
                     /gene="chs1"
                     /note="trans-spliced leader attachment site"
     gene            794. .>8486
                     /gene="chs1"
     mRNA            join(794. .923,1270. .1408,1653. .1795,1853. .2019,2145. .2361,
                     2514. .2683,2865. .3203,3291. .3372,3472. .3635,3680. .3765,
                     3856. .3958,4050. .4203,4263. .4473,4581. .4780,4864. .5040,
                     5115. .5231,5519. .5708,5817. .5915,6372. .6490,6570. .6770,
                     6940. .7110,7256. .7421,7618. .7796,7941. .8010,8326. .>8486)
                     /gene="chs1"
                     /product="chitin synthase"
     CDS             join(795. .923,1270. .1408,1653. .1795,1853. .2019,2145. .2361,
                     2514. .2683,2865. .3203,3291. .3372,3472. .3635,3680. .3765,
                     3856. .3958,4050. .4203,4263. .4473,4581. .4780,4864. .5040,
                     5115. .5231,5519. .5708,5817. .5915,6372. .6490,6570. .6770,
                     6940. .7110,7256. .7421,7618. .7796,7941. .8010,8326. .8486)
                     /gene="chs1"
                     /EC_number="2.4.1.16"
                     /codon_start=1
                     /product="chitin synthase"
                     /protein_id="AAG53086.1"
                     /db_xref="GI:12382256"
                     /translation="MSSSKAKWDSLRSRWNVFRSYSHEYHFNNELTPWMVKTLLILKF
                     LIFVVSHLFLIIGAAISKLVVILLATNIELKSSGSSFEKKCHVGEEKARSTEVTVSVV
                     CALWLIQNIPDVMAIIQSLYRIYKYPKERKGAALTLFMMECSRSMGMAILFFHVFPSL
                     DPLYCLILSVYVVFMPLYLKIKELIHCSFNPIWSLKKRLKLLLHSCSYSVLFLVVFAS
                     CYMWSVSAVPFFKYIALPLGLILSSFGFWDLWINEGHCMSSYHYLYQLKYGTRKMTSM
                     TRFLISALRILTSSAIVWIVKLRYVPFQAVKTSLIEMTDSESTTALYTLAFWIIFLNF
                     VLRFISRLLSAMSMRVLAVLHPLLIVVPLLTMIFIGMCFITPACSIRGYLSIYELRWH
                     CLAWKVSQNSLPEICFGVLWLIIYLCWAYQHVKAPKFSADDDVFDTITTLLNGLTIEQ
                     SLVVYRVSHNKKEIDDNHWSSDEAEVDDGDLRITNDEIDKTVTLYICATMWHETRIEM
                     MQMLKSITKISCFRLDKEHSLMLIERRRSDDIKYRLEAHIFFDDVWEDQMECGRAPNG
                     FFQTFFHLLLELTQNESVNESAENDRVLVNTAYGGRLVVRLPAGTLLFVHLKDKQLVR
                     HKKRWSQVMYLYYLLGHRIMDSHMSVEDRQLEADNTYILAIDGDSKFEPSAVMKLLRL
                     MNAKNEIGCACGRIHPIGEGIMIWYQKFEYAISHWFQKAAEHVFGCVLCAPGCFSLFR
                     ASALMDDNVMHKYTKTASEARHFVQYDQGEDRWLSTLLLKQGYRIEYAAVADAETYAP
                     EGFHEFFNQRRRWTPSSIANTVDLLADYKRVCQSNNSISRMYILYQSMVIGFSLLSPS
                     IVFTMLVYAQVSTFEAESDKMLLYNSLPVLTFIGLCFIADSNSQIIFAKLISVIYGFV
                     MLAVAIATTNQIILETALSPTPIFVLCMIGIFVLAAILHPQEFHNIFYGLIFFLMIPC
                     TYIFMSLYALINLNIINWGTREAASVAIGNNASHTEFRRLLQRFGLHKVIDFVRSTSS
                     WRSAITTTTATTTATTAATDVQEIQKKIDNMKTELQELKDVVLSSSPSTPFKRLVLRE
                     NMNRFLWMDTEYLQVCSRGRLNAAEDEFWNGMIEQYLKPLEFTEEEMRRDVASLVSLR
                     NRIASSVLIVNGLLVLAVFLIQKHKEILSFQYKPYEGFEWSKLSKKTGKFELTDEPLK
                     VEPIGLFIIAFLLIILVVQTCGMFAHRITTLVGAFHEVSSMEDFDFSNKNFENENRIK
                     ENAQQMFDATLYYNAHGADGYIRLHHDTRAACNVLNKLQQSKVMHRSAVSRVSRL"
BASE COUNT     2692 a   1394 c   1487 g   3027 t      1 others
ORIGIN


  Query Match           51.8%;  Score 26.4;  DB 3;  Length 8601;
  Best Local Similarity  75.0%;  Pred. No. 16;
  Matches   33;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       5 acttggtaccgatgaacccactagtaatgaacgaattcatgaac 48
           | ||| ||  |||||   |||||||||||| | ||||||| | |
Db    5389 ATTTGCTATAGATGATTTCACTAGTAATGATCAAATTCATCACC 5346


RESULT   2
AC092410
LOCUS       AC092410   168497 bp    DNA             HTG       04-JUL-2001
DEFINITION  Bos taurus clone RP42-245C6, WORKING DRAFT SEQUENCE, 5 unordered
            pieces.
ACCESSION   AC092410
VERSION     AC092410.1  GI:14595784
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 168497)
  AUTHORS   Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W.,
            Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S., Dietrich,N.L.,
            Granite,S., Guan,X., Gupta,J., Ho,S.-L., Idol,J.R., Karlins,E.,
            Lee-Lin,S.-Q., Legaspi,R., Lim,M., Maduro,Q.L., Maduro,V.B.,
            Maşiello,C., Mastrian,S.D., McCloskey,J.C., McDowell,J.,
            Pearson,R., Prasad,A., Shevchenko,Y., Snyder,B., Stantripop,S.,
            Thomas,J.W., Thomas,P.J., Touchman,J.W., Tsurgeon,C., Vogt,J.L.,
            Walker,M.A., Wetherby,K.D., Zhang,L.-H. and Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 168497)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUL-2001) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     -------------- Genome Center
                Center: NIH Intramural Sequencing Center
                Center code: NISC
                Web site: http://www.nisc.nih.gov
                Contact: nisc_mouse@nhgri.nih.gov
            -------------- Project Information
                Center project name: cjc
                Center clone name: 245C06
            -------------- Summary Statistics
                Sequencing vector: plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.990319
                Consensus quality: 166686 bases at least Q40
                Consensus quality: 167007 bases at least Q30
                Consensus quality: 167085 bases at least Q20
                Insert size: 159000; agarose-fp
                Insert size: 168097; sum-of-contigs
                Quality coverage: 12.17x in Q20 bases; agarose-fp
                Quality coverage: 11.52x in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    12474: contig of 12474 bp in length
            *    12475    12574: gap of unknown length
            *    12575    27577: contig of 15003 bp in length
            *    27578    27677: gap of unknown length
            *    27678    46537: contig of 18860 bp in length
            *    46538    46637: gap of unknown length
            *    46638    80403: contig of 33766 bp in length
            *    80404    80503: gap of unknown length
            *    80504   168497: contig of 87994 bp in length.
FEATURES             Location/Qualifiers
     source          1. .168497
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone="RP42-245C6"
                     /clone_lib="RP42"
     misc_feature    1. .12474
                     /note="assembly_fragment"
     misc_feature    12575. .27577
                     /note="assembly_fragment
                     clone_end:T7
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 6, 2002, 09:15:14 ; Search time 71.9 Seconds
(without alignments)
14.710 Million cell updates/sec

Title: US-09-897-042-16
Perfect score: 262
Sequence: 1 MNLVPMNPLVMNGFCRYPSH.....RLLTKPERRLSWLLPPLSNN 47

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	195.5	74.6	314	1	US-08-040-753-2	Sequence 2, Appli
2	59	22.5	110	1	US-08-081-539-69	Sequence 69, Appl
3	59	22.5	110	1	US-08-466-647-69	Sequence 69, Appl
4	53.5	20.4	124	1	US-08-240-049B-20	Sequence 20, Appl
5	53.5	20.4	124	4	US-08-542-634-22	Sequence 22, Appl
6	53.5	20.4	124	4	US-08-477-292-22	Sequence 22, Appl
7	53.5	20.4	124	5	PCT-US95-13703-22	Sequence 22, Appl
8	52.5	20.0	346	1	US-07-895-300A-1	Sequence 1, Appli
9	52.5	20.0	346	1	US-08-458-367-1	Sequence 1, Appli
10	52.5	20.0	346	5	PCT-US93-05136-1	Sequence 1, Appli
11	52	19.8	923	3	US-08-936-135-6	Sequence 6, Appli
12	51.5	19.7	104	2	US-08-341-843B-35	Sequence 35, Appl
13	51.5	19.7	104	2	US-08-427-497E-40	Sequence 40, Appl
14	51.5	19.7	894	2	US-08-599-455B-2	Sequence 2, Appli
15	51.5	19.7	894	4	US-09-069-781B-2	Sequence 2, Appli
16	51.5	19.7	953	2	US-08-500-857A-2	Sequence 2, Appli
17	51.5	19.7	1162	2	US-08-599-455B-43	Sequence 43, Appl
18	51.5	19.7	1162	4	US-09-069-781B-43	Sequence 43, Appl
19	50	19.1	454	3	US-08-929-329-8	Sequence 8, Appli
20	49.5	18.9	142	2	US-08-392-546C-1	Sequence 1, Appli
21	49.5	18.9	896	2	US-08-640-389A-12	Sequence 12, Appl
22	48.5	18.5	405	1	US-08-351-473B-2	Sequence 2, Appli
23	48.5	18.5	895	4	US-08-827-962-19	Sequence 19, Appl
24	48.5	18.5	895	4	US-08-827-962-21	Sequence 21, Appl
25	48.5	18.5	1162	4	US-08-827-962-15	Sequence 15, Appl
26	48.5	18.5	1162	4	US-08-827-962-20	Sequence 20, Appl
27	48.5	18.5	1162	4	US-08-803-346-1	Sequence 1, Appli
28	48	18.3	373	3	US-08-945-056-2	Sequence 2, Appli
29	48	18.3	379	1	US-08-164-614A-8	Sequence 8, Appli
30	48	18.3	379	2	US-08-456-489B-8	Sequence 8, Appli
31	48	18.3	468	1	US-08-164-614A-7	Sequence 7, Appli
32	48	18.3	468	2	US-08-456-489B-7	Sequence 7, Appli
33	47.5	18.1	120	1	US-08-497-312-26	Sequence 26, Appl
34	47.5	18.1	120	1	US-08-497-312-28	Sequence 28, Appl
35	47.5	18.1	373	3	US-08-911-321-9	Sequence 9, Appli
36	47.5	18.1	373	5	PCT-US95-13975-73	Sequence 73, Appl
37	47.5	18.1	374	1	US-08-220-151-21	Sequence 21, Appl
38	47.5	18.1	374	1	US-08-413-118-21	Sequence 21, Appl
39	47.5	18.1	374	3	US-08-473-446-21	Sequence 21, Appl
40	47	17.9	339	3	US-08-968-563-14	Sequence 14, Appl
41	47	17.9	339	4	US-08-969-683A-14	Sequence 14, Appl
42	47	17.9	375	1	US-08-468-847B-13	Sequence 13, Appl
43	47	17.9	432	4	US-08-702-665A-3	Sequence 3, Appli
44	47	17.9	441	4	US-09-151-102-4	Sequence 4, Appli
45	46.5	17.7	550	1	US-08-484-493-2	Sequence 2, Appli
46	46.5	17.7	550	1	US-08-484-494-2	Sequence 2, Appli
47	46.5	17.7	550	2	US-08-345-212-2	Sequence 2, Appli
48	46.5	17.7	550	4	US-09-249-003-2	Sequence 2, Appli
49	46.5	17.7	591	1	US-08-145-995A-21	Sequence 21, Appl
50	46.5	17.7	591	2	US-08-451-747-21	Sequence 21, Appl

ALIGNMENTS

```
RESULT   1
US-08-040-753-2
; Sequence 2, Application US/08040753
; Patent No. 5464745
;  GENERAL INFORMATION:
;    APPLICANT:  Mierendorf, Robert
;    APPLICANT:  Garber, Richard
;    APPLICANT:  No. 5464745y, Robert
;    APPLICANT:  Hammer, Beth
;    TITLE OF INVENTION:  Protein Ligand Binding
;    TITLE OF INVENTION:  Region Mapping System
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Quarles and Brady
;      STREET:  1 South Pinckney St., Box 2113
;      CITY:  Madison
;      STATE:  WI
;      COUNTRY:  USA
;      ZIP:  53701-2113
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/040,753
;      FILING DATE:  19930331
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Seay, Nicholas J
;      REGISTRATION NUMBER:  27386
;      REFERENCE/DOCKET NUMBER:  70-399-9001-1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  608-251-5000
;      TELEFAX:  608-251-9166
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  314 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-040-753-2
```

```
  Query Match             74.6%;  Score 195.5;  DB 1;  Length 314;
  Best Local Similarity   76.5%;  Pred. No. 6.4e-18;
  Matches   39;  Conservative    3;  Mismatches    4;  Indels    5;  Gaps    2;

Qy       2 NLVP-MNPLVMNG----FCRYPSHWRPLEQIRLLTKPERRLSWLLPPLSNN 47
           :|||  :|||       ||||||||||||||||||||:||||||||||||
Db     264 SLVPSSDPLVTAASVLEFCRYPSHWRPLEQIRLLTKPERKLSWLLPPLSNN 314


RESULT   2
US-08-081-539-69
; Sequence 69, Application US/08081539
; Patent No. 5501962
;  GENERAL INFORMATION:
;    APPLICANT:  Braford-Goldberg, Sarah R.
;    APPLICANT:  Easton, Alan M.
;    APPLICANT:  Klein, Barbara K.
;    APPLICANT:  McKearn, John P.
;    APPLICANT:  Olins, Peter O.
;    TITLE OF INVENTION:  Interleukin-3 (IL-3) Human/Murine
;    TITLE OF INVENTION:  Chimeric Hybrid Polypeptides
;    NUMBER OF SEQUENCES:  121
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Mary Jo Kanady, G. D. Searle & Co., Corporate
;      ADDRESSEE:  Patent Dept.
;      STREET:  P. O. Box 5110
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60680
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/081,539
;      FILING DATE:  19930621
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kanady, Mary J.
;      REGISTRATION NUMBER:  28623
;      REFERENCE/DOCKET NUMBER:  2724
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (708)470-6501
;      TELEFAX:  (708)470-6881
;  INFORMATION FOR SEQ ID NO:  69:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  110 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-081-539-69


  Query Match             22.5%;  Score 59;  DB 1;  Length 110;
  Best Local Similarity   33.3%;  Pred. No. 1.1;
  Matches   15;  Conservative    8;  Mismatches   16;  Indels    6;  Gaps    1;

Qy       2 NLVPMNPLVMNGFCRYPSHWRPLEQIRLLTKPERRLSWLLPPLSN 46
           ||:|  ||      |:| | : |:: |      |:|:: |  | |
Db      67 NLLPCLPLATAAPTRHPIHIKDLDEFR------RKLTFYLKTLEN 105


RESULT   3
US-08-466-647-69
; Sequence 69, Application US/08466647
; Patent No. 5543141
;  GENERAL INFORMATION:
;    APPLICANT:  Braford-Goldberg, Sarah R.
;    APPLICANT:  Easton, Alan M.
;    APPLICANT:  Klein, Barbara K.
;    APPLICANT:  McKearn, John P.
;    APPLICANT:  Olins, Peter O.
;    TITLE OF INVENTION:  Interleukin-3 (IL-3) Human/Murine
;    TITLE OF INVENTION:  Chimeric Hybrid Polypeptides
;    NUMBER OF SEQUENCES:  121
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Mary Jo Kanady, G. D. Searle & Co., Corporate
;      ADDRESSEE:  Patent Dept.
;      STREET:  P. O. Box 5110
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60680
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/466,647
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/081,539
;      FILING DATE:  21-JUN-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kanady, Mary J.
;      REGISTRATION NUMBER:  28623
;      REFERENCE/DOCKET NUMBER:  2724
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (708)470-6501
;      TELEFAX:  (708)470-6881
;  INFORMATION FOR SEQ ID NO:  69:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  110 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-466-647-69


  Query Match             22.5%;  Score 59;  DB 1;  Length 110;
  Best Local Similarity   33.3%;  Pred. No. 1.1;
  Matches   15;  Conservative    8;  Mismatches   16;  Indels    6;  Gaps    1;

Qy       2 NLVPMNPLVMNGFCRYPSHWRPLEQIRLLTKPERRLSWLLPPLSN 46
           ||:|  ||      |:| | : |:: |      |:|:: |  | |
Db      67 NLLPCLPLATAAPTRHPIHIKDLDEFR------RKLTFYLKTLEN 105


RESULT   4
US-08-240-049B-20
; Sequence 20, Application US/08240049B
; Patent No. 5686239
;  GENERAL INFORMATION:
;    APPLICANT:  Reyes, Gregory R.
;    APPLICANT:  Tam, Albert W.
;    APPLICANT:  Yarbough, Patrice O.
;    TITLE OF INVENTION:  Hepatitis E Virus Peptides and Methods
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genelabs Technologies, Inc.
;      STREET:  505 Penobscot Drive
;      CITY:  Redwood City
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94063
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/240,049B
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 6, 2002, 09:18:42 ; Search time 103.34 Seconds
(without alignments)
33.689 Million cell updates/sec

Title: US-09-897-042-16
Perfect score: 262
Sequence: 1 MNLVPMNPLVMNGFCRYPSH.....RLLTKPERRLSWLLPPLSNN 47

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : A_Geneseq_1101:*
1: /SIDS2/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	195.5	74.6	314	16	AAR87027	T7 gene 10 leader
2	84	32.1	654	22	AAG83278	Chlamydia trachoma
3	80.5	30.7	691	22	AAG83271	Chlamydia trachoma
4	80	30.5	683	22	AAG83282	Chlamydia trachoma
5	79	30.2	646	22	AAG83272	Chlamydia trachoma
6	77.5	29.6	715	22	AAG83273	Chlamydia trachoma
7	77	29.4	715	22	AAG83275	Chlamydia trachoma
8	75	28.6	583	22	AAG83281	Chlamydia trachoma
9	74.5	28.4	518	22	AAG83276	Chlamydia trachoma
10	74	28.2	585	22	AAG83277	Chlamydia trachoma
11	74	28.2	700	22	AAG83279	Chlamydia trachoma
12	73	27.9	487	22	AAG83280	Chlamydia trachoma
13	73	27.9	619	22	AAG83270	Chlamydia trachoma
14	73	27.9	631	22	AAG83274	Chlamydia trachoma
15	70	26.7	666	20	AAY22471	Human C-term mychi
16	70	26.7	666	22	AAB59203	C-terminal mychis
17	70	26.7	669	20	AAY22470	Human Cterm V5 his
18	70	26.7	669	22	AAB59202	C-terminal V5 His
19	59	22.5	110	16	AAR69757	pMON5968 (Met-15-1
20	59	22.5	132	22	AAM40409	Human polypeptide
21	57	21.8	442	22	AAM41638	Human polypeptide
22	56	21.4	281	22	AAB48050	Signal transductio
23	56	21.4	391	22	AAB87488	B thuringiensis 44
24	55.5	21.2	151	19	AAW52130	Putative insectici
25	54.5	20.8	553	20	AAY38796	Neisseria meningit
26	54.5	20.8	558	20	AAY38797	Neisseria meningit
27	53.5	20.4	123	14	AAR38786	HEV ORF3 protein.
28	53.5	20.4	123	14	AAR39307	Mexico strain HEV
29	53.5	20.4	123	17	AAR96098	Hepatitis E virus
30	53.5	20.4	123	20	AAW93398	Human HEV ORF 3 pr
31	53.5	20.4	124	18	AAW35829	Hepatitis E virus
32	53	20.2	132	9	AAP80381	Interleukin-3 vari
33	53	20.2	133	9	AAP80383	Interleukin-3 vari
34	53	20.2	155	21	AAB11515	SEN virus protein
35	53	20.2	753	21	AAB11525	SEN virus protein
36	53	20.2	1215	21	AAY55791	Superheat-resistan
37	53	20.2	1215	22	AAB84784	Pyrococcus heat re
38	52.5	20.0	51	22	AAM17391	Peptide #3825 enco
39	52.5	20.0	51	22	AAM29911	Peptide #3948 enco
40	52.5	20.0	51	22	AAM05079	Peptide #3761 enco
41	52.5	20.0	260	17	AAW14408	Human DNase I vari
42	52.5	20.0	260	17	AAW14409	Human DNase I vari
43	52.5	20.0	260	17	AAW14410	Human DNase I vari
44	52.5	20.0	260	17	AAW14411	Human DNase I vari
45	52.5	20.0	260	17	AAW14412	Human DNase I vari
46	52.5	20.0	260	17	AAW14413	Human DNase I vari
47	52.5	20.0	260	17	AAW14414	Human DNase I vari
48	52.5	20.0	260	17	AAW14415	Human DNase I vari
49	52.5	20.0	260	17	AAW14416	Human DNase I vari
50	52.5	20.0	260	17	AAW14417	Human DNase I vari

ALIGNMENTS

RESULT 1
AAR87027
ID AAR87027 standard; Protein; 314 AA.
XX
AC AAR87027;
XX
DT 18-MAR-1996 (first entry)
XX
DE T7 gene 10 leader sequence product.
XX
KW Plasmid pTOPE-1b(+); vector; ligand binding domain; epitope mapping;
KW antigen; Escherichia coli.
XX
OS Bacteriophage T7.
XX
PN US5464745-A.
XX
PD 07-NOV-1995.
XX
PF 31-MAR-1993; 93US-0040753.
XX
PR 31-MAR-1993; 93US-0040753.
XX
PA (NOVA-) NOVAGEN INC.
XX
PI Garber R, Hammer B, Mierendorf R, Novy R;
XX
DR WPI; 1995-392610/50.
DR N-PSDB; AAT07310.

```
XX
PT   Mapping ligand binding domains, esp. epitope(s), of proteins - by
PT   expressing peptide(s) encoded by random gene fragments and testing
PT   for ligand binding
XX
PS   Disclosure; Column 15-18; 12pp; English.
XX
CC   The phage T7 gene 10 leader sequence product (AAR87027) is encoded
CC   by prokaryotic expression vector pTOPE-1b(+) (AAT07310). A fusion
CC   of the leader sequence and a putative ligand-binding domain (LBD),
CC   esp. antigenic region, of a protein is obtd. following insertion
CC   of putative LBD-encoding DNA into the vector. The fusion protein
CC   accumulates as inclusion bodies in Escherichia coli host cells and
CC   can be screened for its ability to bind a ligand.
XX
SQ   Sequence   314 AA;


  Query Match             74.6%;  Score 195.5;  DB 16;  Length 314;
  Best Local Similarity   76.5%;  Pred. No. 4.9e-17;
  Matches   39;  Conservative    3;  Mismatches    4;  Indels    5;  Gaps    2;

Qy       2 NLVP-MNPLVMNG----FCRYPSHWRPLEQIRLLTKPERRLSWLLPPLSNN 47
           :|||  :|||       |||||||||||||||||||||:||||||||||
Db     264 slvpssdplvtaasvlefcrypshwrpleqirlltkperklswllpplsnn 314


RESULT   2
AAG83278
ID   AAG83278 standard; Protein; 654 AA.
XX
AC   AAG83278;
XX
DT   05-SEP-2001  (first entry)
XX
DE   Chlamydia trachomatis PmpB(3) fusion protein.
XX
KW   Chlamydia; vaccine; infection; fusion protein; antigen;
KW   pelvic inflammatory disease; trachoma; atherosclerosis; heart disease;
KW   acute respiratory tract infection; Cap1; CT529; OMCB;
KW   polymorphic membrane protein; pmp; thiol specific antioxidant; TSA.
XX
OS   Chlamydia trachomatis.
XX
PN   WO200140474-A2.
XX
PD   07-JUN-2001.
XX
PF   04-DEC-2000; 2000WO-US32919.
XX
PR   03-DEC-1999;   99US-0454684.
PR   19-APR-2000; 2000US-0556877.
PR   20-JUN-2000; 2000US-0598419.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Probst P,  Bhatia A,  Skeiky YAW,  Fling SP,  Scholler J;
XX
DR   WPI; 2001-374831/39.
XX
PT   Chlamydia polypeptides and fusion proteins useful for preventing pelvic
PT   inflammatory disease, trachoma, acute respiratory tract infections,
PT   atherosclerosis and heart disease -
XX
PS   Claim 70; Page 284-285; 295pp; English.
XX
CC   The present sequence is provided in a specification relating to
CC   compounds and methods for the treatment and diagnosis of chlamydial
CC   infection. The compounds provided include polypeptides and fusion
CC   proteins comprising immunogenic portions of Chlamydia antigens
CC   and DNA sequences encoding such polypeptides. They are useful for
CC   vaccinating against chlamydial infection, which causes pelvic
CC   inflammatory disease, trachoma, acute respiratory tract infections,
CC   atherosclerosis and heart disease.
XX
SQ   Sequence   654 AA;


  Query Match             32.1%;  Score 84;  DB 22;  Length 654;
  Best Local Similarity   75.0%;  Pred. No. 0.018;
  Matches   12;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      14 FCRYPSHWRPLEQIRL 29
           ||||||||||:|: :
Db     137 fcrypshwrpldqlni 152


RESULT   3
AAG83271
ID   AAG83271 standard; Protein; 691 AA.
XX
AC   AAG83271;
XX
DT   05-SEP-2001  (first entry)
XX
DE   Chlamydia trachomatis PmpA(C-term) fusion protein.
XX
KW   Chlamydia; vaccine; infection; fusion protein; antigen;
KW   pelvic inflammatory disease; trachoma; atherosclerosis; heart disease;
KW   acute respiratory tract infection; Cap1; CT529; OMCB;
KW   polymorphic membrane protein; pmp; thiol specific antioxidant; TSA.
XX
OS   Chlamydia trachomatis.
XX
PN   WO200140474-A2.
XX
PD   07-JUN-2001.
XX
PF   04-DEC-2000; 2000WO-US32919.
XX
PR   03-DEC-1999;   99US-0454684.
PR   19-APR-2000; 2000US-0556877.
PR   20-JUN-2000; 2000US-0598419.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Probst P,  Bhatia A,  Skeiky YAW,  Fling SP,  Scholler J;
XX
DR   WPI; 2001-374831/39.
XX
PT   Chlamydia polypeptides and fusion proteins useful for preventing pelvic
PT   inflammatory disease, trachoma, acute respiratory tract infections,
PT   atherosclerosis and heart disease -
XX
PS   Claim 70; Page 267-268; 295pp; English.
XX
CC   The present sequence is provided in a specification relating to
CC   compounds and methods for the treatment and diagnosis of chlamydial
CC   infection. The compounds provided include polypeptides and fusion
CC   proteins comprising immunogenic portions of Chlamydia antigens
CC   and DNA sequences encoding such polypeptides. They are useful for
CC   vaccinating against chlamydial infection, which causes pelvic
CC   inflammatory disease, trachoma, acute respiratory tract infections,
CC   atherosclerosis and heart disease.
XX
SQ   Sequence   691 AA;


  Query Match             30.7%;  Score 80.5;  DB 22;  Length 691;
  Best Local Similarity   50.0%;  Pred. No. 0.053;
  Matches   17;  Conservative    1;  Mismatches   15;  Indels    1;  Gaps    1;

Qy      14 FCRYPSHWRPLEQIRLLTKPERRLSWLLPPLSNN 47
           ||||||||||  |     |   ||    |: |
Db     137 fcrypshwrplhsihff-dpvmalsassspiqin 169
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 6, 2002, 09:33:18 ; Search time 135.97 Seconds
(without alignments)
50.561 Million cell updates/sec

Title: US-09-897-042-16
Perfect score: 262
Sequence: 1 MNLVPMNPLVMNGFCRYPSH.....RLLTKPERRLSWLLPPLSNN 47

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : SPTREMBL_17:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	114.5	43.7	90	2	Q47336	Q47336 escherichia
2	83.5	31.9	217	2	O53022	O53022 escherichia
3	68	26.0	2081	1	Q9HJ37	Q9hj37 thermoplasm
4	60.5	23.1	132	4	Q9NWS2	Q9nws2 homo sapien
5	59	22.5	132	4	Q9BVV8	Q9bvv8 homo sapien
6	59	22.5	215	11	Q9CS38	Q9cs38 mus musculu
7	59	22.5	216	11	Q9DAY6	Q9day6 mus musculu
8	59	22.5	216	11	Q9D064	Q9d064 mus musculu
9	59	22.5	216	11	Q9D053	Q9d053 mus musculu
10	59	22.5	216	11	Q9D033	Q9d033 mus musculu
11	58	22.1	395	12	O56762	O56762 dengue viru
12	58	22.1	396	12	O10247	O10247 dengue viru
13	58	22.1	399	12	O10248	O10248 dengue viru
14	58	22.1	399	12	O10254	O10254 dengue viru
15	58	22.1	400	12	O10255	O10255 dengue viru
16	58	22.1	483	10	Q9FJ55	Q9fj55 arabidopsis
17	57	21.8	420	4	O14979	O14979 homo sapien
18	55.5	21.2	281	1	Q9V217	Q9v217 pyrococcus
19	55.5	21.2	311	9	Q9MCC8	Q9mcc8 lactobacill
20	55	21.0	1327	5	Q9VLN6	Q9vln6 drosophila
21	54.5	20.8	279	1	O58021	O58021 pyrococcus
22	54.5	20.8	362	2	Q9K9V8	Q9k9v8 bacillus ha
23	54.5	20.8	553	2	Q9JXN7	Q9jxn7 neisseria m
24	54.5	20.8	558	2	Q9JW86	Q9jw86 neisseria m
25	54.5	20.8	699	10	Q9AVC8	Q9avc8 pisum sativ
26	54	20.6	123	2	O52893	O52893 coxiella bu
27	54	20.6	152	2	Q45954	Q45954 coxiella bu
28	54	20.6	226	2	O25822	O25822 helicobacte
29	54	20.6	378	11	Q9D427	Q9d427 mus musculu
30	54	20.6	387	12	O10251	O10251 dengue viru
31	54	20.6	581	11	O35120	O35120 mus musculu
32	54	20.6	581	11	Q9QYT7	Q9qyt7 mus musculu
33	54	20.6	612	8	Q9G6Q3	Q9g6q3 diplophos t
34	53.5	20.4	185	8	Q33889	Q33889 adela trigr
35	53	20.2	332	2	Q51888	Q51888 prevotella
36	53	20.2	380	2	O68873	O68873 myxococcus
37	53	20.2	902	12	O71101	O71101 bovine aden
38	53	20.2	1215	1	Q9UWR7	Q9uwr7 pyrococcus
39	53	20.2	7829	5	Q18559	Q18559 caenorhabdi
40	52.5	20.0	227	8	Q9T7A8	Q9t7a8 anopheles m
41	52.5	20.0	452	2	Q53659	Q53659 streptomyce
42	52.5	20.0	4845	11	O88738	O88738 mus musculu
43	52	19.8	411	4	O60643	O60643 homo sapien
44	52	19.8	463	10	Q9LJ48	Q9lj48 arabidopsis
45	52	19.8	528	4	Q9NQ62	Q9nq62 homo sapien
46	52	19.8	534	4	Q9NQ61	Q9nq61 homo sapien
47	52	19.8	757	5	Q9VLI5	Q9vli5 drosophila
48	52	19.8	805	5	Q9U3Z6	Q9u3z6 drosophila
49	52	19.8	837	2	Q9HTF2	Q9htf2 pseudomonas
50	52	19.8	921	11	Q9QX38	Q9qx38 rattus norv

ALIGNMENTS

RESULT 1
Q47336
ID Q47336 PRELIMINARY; PRT; 90 AA.
AC Q47336;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE LACZ-ALPHA PEPTIDE.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RA Maneewannakul K., Maneewannakul S., Ippen-Ihler K.;
RL Submitted (AUG-1991) to the EMBL/GenBank/DDBJ databases.
DR EMBL; M74750; AAA24056.1; -.
DR HSSP; P00722; 1BGL.
SQ SEQUENCE 90 AA; 10171 MW; 822AF741095111DC CRC64;

Query Match 43.7%; Score 114.5; DB 2; Length 90;
Best Local Similarity 48.3%; Pred. No. 8.8e-08;
Matches 28; Conservative 5; Mismatches 8; Indels 17; Gaps 3;

```
Qy       7 NPLV--MNGFCRYP--SHWRPLEQ------------IRLLTKPERRLSWLLPPLSNN 47
           || |  :|    :|  : ||  |:             |||||||||:||||||||||
Db      33 NPGVTQLNRLAAHPPFASWRNSEEARTDRPSQQLRSLIRLLTKPERKLSWLLPPLSNN 90
```

RESULT 2
O53022
ID O53022 PRELIMINARY; PRT; 217 AA.
AC O53022;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)

```
DE   CLONING VECTOR PZERO-2T.
GN   FUSED-CCDB.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98172738; PubMed=9511747;
RA   Gabant P., Szpirer C.Y., Couturier M., Faelan M.;
RT   "Direct selection cloning vectors adapted to the genetic analysis of
RT   gram-negative bacteria and their plasmids.";
RL   Gene 207:87-92(1998).
DR   EMBL; Y10545; CAA71575.1; -.
DR   HSSP; P05703; 3VUB.
DR   InterPro; IPR002712; CcdB.
DR   Pfam; PF01845; CcdB; 1.
DR   ProDom; PD012578; CcdB; 1.
SQ   SEQUENCE   217 AA;  24893 MW;  8DCFF1B864228035 CRC64;


  Query Match             31.9%;  Score 83.5;  DB 2;  Length 217;
  Best Local Similarity   62.1%;  Pred. No. 0.0027;
  Matches   18;  Conservative    2;  Mismatches    4;  Indels    5;  Gaps    2;

Qy       2 NLVP-MNPLVMNG----FCRYPSHWRPLE 25
           :|||  :|||       ||||||||||||
Db      21 SLVPSSDPLVTAASVLEFCRYPSHWRPLE 49


RESULT   3
Q9HJ37
ID   Q9HJ37      PRELIMINARY;      PRT;  2081 AA.
AC   Q9HJ37;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   CONSERVED HYPOTHETICAL MEMBRANE PROTEIN.
GN   TA1136.
OS   Thermoplasma acidophilum.
OC   Archaea; Euryarchaeota; Thermoplasmales; Thermoplasmaceae;
OC   Thermoplasma.
OX   NCBI_TaxID=2303;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DSM 1728;
RX   MEDLINE=20479972; PubMed=11029001;
RA   Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
RA   Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;
RT   "The genome sequence of the thermoacidophilic scavenger Thermoplasma
RT   acidophilum.";
RL   Nature 407:508-513(2000).
DR   EMBL; AL445066; CAC12262.1; -.
KW   Complete proteome.
SQ   SEQUENCE   2081 AA;  227496 MW;  6E02AA6470DB2E8D CRC64;


  Query Match             26.0%;  Score 68;  DB 1;  Length 2081;
  Best Local Similarity   38.6%;  Pred. No. 3.1;
  Matches   17;  Conservative    7;  Mismatches   16;  Indels    4;  Gaps    2;

Qy       4 VPMNPLVMNGFCRYPSHWR--PLEQIRLLTKPERRLSWLLPPLS 45
           || ||   |    :|| |:  |:::   ||   :  : ||||||
Db     850 VPYNP--YNNTTAHPSAWKIVPIQEAYTLTSEGKGKAQLLPPLS 891


RESULT   4
Q9NWS2
ID   Q9NWS2      PRELIMINARY;      PRT;  132 AA.
AC   Q9NWS2;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   CDNA FLJ20640 FIS, CLONE KAT02911.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
RA   Suzuki Y., Obayashi M., Nishi T., Shibahara T., Tanaka T.,
RA   Nakamura Y., Isogai T., Sugano S.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK000647; BAA91305.1; -.
SQ   SEQUENCE   132 AA;  14013 MW;  D9E8CB76D4920AF2 CRC64;


  Query Match             23.1%;  Score 60.5;  DB 4;  Length 132;
  Best Local Similarity   37.5%;  Pred. No. 1.8;
  Matches   15;  Conservative    5;  Mismatches   19;  Indels    1;  Gaps    1;

Qy       4 VPMNPLVMNGFCRYPSHW-RPLEQIRLLTKPERRLSWLLP 42
           || |||:   |  :|  |  |:    || :| :    |||
Db      12 VPSNPLLHRSFPCWPRGWSHPVPTRELLLEPAQPADLLLP 51


RESULT   5
Q9BVV8
ID   Q9BVV8      PRELIMINARY;      PRT;  132 AA.
AC   Q9BVV8;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   HYPOTHETICAL 14.0 KDA PROTEIN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=CERVIX CARCINOMA;
RA   Strausberg R.;
RL   Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC000890; AAH00890.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   132 AA;  13997 MW;  B5E8DF6395C61FB3 CRC64;


  Query Match             22.5%;  Score 59;  DB 4;  Length 132;
  Best Local Similarity   39.0%;  Pred. No. 2.8;
  Matches   16;  Conservative    5;  Mismatches   18;  Indels    2;  Gaps    2;

Qy       4 VPMNPLVMNGFCRYPSHW-RPLEQIRLLTKPERRLSWLLPP 43
           || |||:   |  :|  |  |:    || :| :    ||||
Db      12 VPSNPLLHRSFPCWPRGWSHPVPTRELLLEPAQPAD-LLPP 51


RESULT   6
Q9CS38
ID   Q9CS38      PRELIMINARY;      PRT;  215 AA.
AC   Q9CS38;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   1600012F09RIK PROTEIN (FRAGMENT).
GN   1600012F09RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=EMBRYO;
RX   MEDLINE=21085660; PubMed=11217851;
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 6, 2002, 09:34:40 ; Search time 68.88 Seconds
(without alignments)
25.018 Million cell updates/sec

Title: US-09-897-042-16
Perfect score: 262
Sequence: 1 MNLVPMNPLVMNGFCRYPSH.....RLLTKPERRLSWLLPPLSNN 47

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	55	21.0	421	1	CRT2_BOVIN	P42918 bos taurus
2	55	21.0	455	1	HMC2_BLAGE	P54870 blattella g
3	54	20.6	970	1	T3RE_BPP1	P08764 bacteriopha
4	53.5	20.4	123	1	VST1_HEVME	Q03499 hepatitis e
5	53.5	20.4	255	1	RCEM_RHOPH	P51751 rhodospiril
6	53	20.2	309	1	AGAR_STRCO	P07883 streptomyce
7	53	20.2	911	1	HEX_ADEB3	P03278 bovine aden
8	52.5	20.0	282	1	DRN1_HUMAN	P24855 homo sapien
9	52	19.8	400	1	CRT1_BOVIN	P52193 bos taurus
10	52	19.8	416	1	CRTC_MOUSE	P14211 mus musculu
11	52	19.8	416	1	CRTC_RAT	P18418 rattus norv
12	52	19.8	417	1	CRTC_HUMAN	P27797 homo sapien
13	52	19.8	418	1	CRTC_RABIT	P15253 oryctolagus
14	52	19.8	922	1	NRP1_RAT	Q9qwj9 rattus norv
15	52	19.8	923	1	NRP1_MOUSE	P97333 mus musculu
16	52	19.8	1396	1	VCAP_VZVD	P09245 varicella-z
17	51.5	19.7	375	1	RND_ECOLI	P09155 escherichia
18	51.5	19.7	953	1	PODK_FLABI	Q39735 flaveria bi
19	51.5	19.7	953	1	PODK_FLATR	P22221 flaveria tr
20	51.5	19.7	1162	1	LEPR_MOUSE	P48356 mus musculu
21	51.5	19.7	1266	1	NGCA_CHICK	Q03696 gallus gall
22	51.5	19.7	4829	1	BIR6_HUMAN	Q9nr09 homo sapien
23	51	19.5	278	1	RHAS_ECOLI	P09377 escherichia
24	51	19.5	923	1	NRP1_HUMAN	O14786 homo sapien
25	50.5	19.3	184	1	YCDY_ECOLI	P75915 escherichia
26	50.5	19.3	277	1	RHAS_SALTY	P27029 salmonella
27	50.5	19.3	2182	1	CABI_RAT	O88480 rattus norv
28	50	19.1	589	1	MDLA_BUCAI	P57551 buchnera ap
29	50	19.1	607	1	LPXB_CHLTR	O84416 chlamydia t
30	49.5	18.9	322	1	CRJB_TRICY	Q03443 tripedalia
31	49.5	18.9	359	1	CKR3_MOUSE	P51678 mus musculu
32	49.5	18.9	405	1	B3AR_CAPHI	Q9xt57 capra hircu
33	49.5	18.9	469	1	TRME_SYNEL	Q9x9t0 synechococc
34	49.5	18.9	644	1	YEAG_ECOLI	P77391 escherichia
35	49.5	18.9	985	1	AGLU_ASPOR	Q12558 aspergillus
36	49	18.7	610	1	PTMA_BACSU	P42956 bacillus su
37	49	18.7	2054	1	YCF2_PINTH	P41653 pinus thunb
38	48.5	18.5	228	1	COX2_CULQU	P50693 culex quinq
39	48.5	18.5	285	1	PLCA_MOUSE	O35083 mus musculu
40	48.5	18.5	297	1	MT04_AQUAE	O66479 aquifex aeo
41	48.5	18.5	405	1	B3AR_BOVIN	P46626 bos taurus
42	48.5	18.5	446	1	AP50_SCHPO	Q09718 schizosacch
43	48.5	18.5	493	1	XYLB_HAEIN	P44401 haemophilus
44	48.5	18.5	587	1	TALA_BFDV	P13894 budgerigar
45	48.5	18.5	1026	1	EX5B_CHLTR	O84645 chlamydia t
46	48.5	18.5	1162	1	LEPR_RAT	Q62959 rattus norv
47	48	18.3	331	1	PLIB_AGKBL	O93233 agkistrodon
48	48	18.3	468	1	IL9R_MOUSE	Q01114 mus musculu
49	48	18.3	504	1	Y795_METJA	Q58205 methanococc
50	48	18.3	532	1	PUR2_ARATH	P52420 arabidopsis

ALIGNMENTS

```
RESULT   1
CRT2_BOVIN
ID   CRT2_BOVIN     STANDARD;      PRT;   421 AA.
AC   P42918;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   CALRETICULIN, BRAIN ISOFORM 2 PRECURSOR (CRP55) (CALREGULIN) (HACBP).
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=93385184; PubMed=8373827;
RA   Liu N., Fine R.E., Johnson R.J.;
RT   "Comparison of cDNAs from bovine brain coding for two isoforms of
RT   calreticulin.";
RL   Biochim. Biophys. Acta 1202:70-76(1993).
CC   -!- FUNCTION: THIS PROTEIN BINDS CALCIUM. THERE ARE BOTH HIGH AND
CC       LOW AFFINITY CALCIUM-BINDING SITES.
CC   -!- SUBUNIT: MONOMER (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: ENDOPLASMIC RETICULUM LUMEN.
CC   -!- SIMILARITY: BELONGS TO THE CALRETICULIN FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L13462; AAC37307.1; -.
DR   InterPro; IPR001580; Calreticulin.
DR   InterPro; IPR000886; ER_target.
DR   Pfam; PF00262; calreticulin; 1.
DR   PRINTS; PR00626; CALRETICULIN.
DR   ProDom; PD001866; Calreticulin; 1.
DR   PROSITE; PS00014; ER_TARGET; 1.
DR   PROSITE; PS00803; CALRETICULIN_1; 1.
DR   PROSITE; PS00804; CALRETICULIN_2; 1.
DR   PROSITE; PS00805; CALRETICULIN_REPEAT; 3.
KW   Endoplasmic reticulum; Calcium-binding; Repeat; Signal.
FT   SIGNAL        1     34       POTENTIAL.
FT   CHAIN        35    421       CALRETICULIN, BRAIN ISOFORM 2.
FT   DOMAIN       35    201       N-DOMAIN.
FT   DOMAIN      202    312       P-DOMAIN.
FT   DOMAIN      313    421       C-DOMAIN.
```

```
FT   DOMAIN       195    259       4 X APPROXIMATE REPEATS.
FT   REPEAT       195    206       1-1.
FT   REPEAT       214    225       1-2.
FT   REPEAT       231    242       1-3.
FT   REPEAT       248    259       1-4.
FT   DOMAIN       263    301       3 X APPROXIMATE REPEATS.
FT   REPEAT       263    273       2-1.
FT   REPEAT       277    287       2-2.
FT   REPEAT       291    301       2-3.
FT   DOMAIN       366    411       ASP/GLU/LYS-RICH.
FT   DISULFID     141    167       BY SIMILARITY.
FT   CARBOHYD     183    183       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   SITE         418    421       PREVENT SECRETION FROM ER.
SQ   SEQUENCE   421 AA;  48812 MW;  0257E959F71528BC CRC64;


  Query Match             21.0%;  Score 55;  DB 1;  Length 421;
  Best Local Similarity   30.8%;  Pred. No. 12;
  Matches   12;  Conservative    6;  Mismatches   15;  Indels    6;  Gaps    2;

Qy       8 PLVMNGFCRYPSHWRPLEQIRLLTKPERRLSWLLPPLSN 46
           ||: |    |   |:|    | :  || :  |: | : |
Db     268 PLIQNP--EYKGEWKP----RQIDNPEYKGIWIHPEIDN 300


RESULT   2
HMC2_BLAGE
ID   HMC2_BLAGE     STANDARD;      PRT;   455 AA.
AC   P54870;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   HYDROXYMETHYLGLUTARYL-COA SYNTHASE 2 (EC 4.1.3.5) (HMG-COA SYNTHASE 2)
DE   (3-HYDROXY-3-METHYLGLUTARYL COENZYME A SYNTHASE 2).
GN   HMGCS-2.
OS   Blattella germanica (German cockroach).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattaria;
OC   Blaberoidea; Blattellidae; Blattellinae; Blattella.
OX   NCBI_TaxID=6973;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94216267; PubMed=7909314;
RA   Buesa C., Martinez-Gonzalez J., Casals N., Haro D., Piulachs M.D.,
RA   Belles X., Hegardt F.G.;
RT   "Blattella germanica has two HMG-CoA synthase genes. Both are
RT   regulated in the ovary during the gonadotrophic cycle.";
RL   J. Biol. Chem. 269:11707-11713(1994).
CC   -!- FUNCTION: THIS ENZYME CONDENSES ACETYL-COA WITH ACETOACETYL-COA TO
CC       FORM HMG-COA, WHICH IS THE SUBSTRATE FOR HMG-COA REDUCTASE (BY
CC       SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: (S)-3-HYDROXY-3-METHYLGLUTARYL-COA + COA =
CC       ACETYL-COA + H(2)O + ACETOACETYL-COA.
CC   -!- PATHWAY: PRODUCTION OF MEVALONATE FROM HMG-COA PRIOR TO THE
CC       SYNTHESIS OF STEROLS AND ISOPRENOIDS SUCH AS JUVENILE HORMONES.
CC   -!- SIMILARITY: BELONGS TO THE HMG-COA SYNTHASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X77516; CAA54652.1; -.
DR   InterPro; IPR000590; HMG_CoA_synt.
DR   Pfam; PF01154; HMG_CoA_synt; 1.
DR   PROSITE; PS01226; HMG_COA_SYNTHASE; 1.
KW   Lyase; Sterol biosynthesis; Multigene family.
FT   ACT_SITE     120    120       POTENTIAL.
SQ   SEQUENCE   455 AA;  51425 MW;  DD030356548A63B1 CRC64;


  Query Match             21.0%;  Score 55;  DB 1;  Length 455;
  Best Local Similarity   45.5%;  Pred. No. 13;
  Matches   10;  Conservative    6;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       9 LVMNGFCRYPSHWRPLEQIRLL 30
           ||:| | :||  :: |:|:| |
Db     269 LVLNDFIQYPEKYQDLQQLRNL 290


RESULT   3
T3RE_BPP1
ID   T3RE_BPP1      STANDARD;      PRT;   970 AA.
AC   P08764;
DT   01-NOV-1988 (Rel. 09, Created)
DT   01-NOV-1988 (Rel. 09, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   TYPE III RESTRICTION-MODIFICATION SYSTEM ECOPI ENZYME RES
DE   (EC 3.1.21.5).
GN   RES.
OS   Bacteriophage P1.
OC   Viruses; dsDNA viruses, no RNA stage; Tailed phages; Myoviridae.
OX   NCBI_TaxID=10678;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88245189; PubMed=2837577;
RA   Huembelin M., Suri B., Rao D.N., Hornby D.P., Eberle H., Pripfl T.,
RA   Kenel S., Bickle T.A.;
RT   "Type III DNA restriction and modification systems EcoP1 and EcoP15.
RT   Nucleotide sequence of the EcoP1 operon, the EcoP15 mod gene and some
RT   EcoP1 mod mutants.";
RL   J. Mol. Biol. 200:23-29(1988).
CC   -!- FUNCTION: THIS PROTEIN CUTS THE DNA SOME 25 BASE-PAIRS TO THE
CC       3'-END OF THE RECOGNITION SITE. IT IS ONLY REQUIRED FOR
CC       RESTRICTION BUT NEEDS THE PRESENCE OF THE MODIFICATION ENZYME.
CC   -!- COFACTOR: MAGNESIUM AND ATP.
CC   -!- SUBUNIT: CONTAINS TWO DIFFERENT SUBUNITS: RES AND MOD.
CC   -!- SIMILARITY: WITH OTHER TYPE III RES PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X06287; CAA29615.1; -.
DR   PIR; S01352; S01352.
DR   REBASE; 988; EcoPI.
KW   Hydrolase; Nuclease; Endonuclease; Restriction system; Helicase.
SQ   SEQUENCE   970 AA;  111458 MW;  B599110154D723AA CRC64;


  Query Match             20.6%;  Score 54;  DB 1;  Length 970;
  Best Local Similarity   48.3%;  Pred. No. 38;
  Matches   14;  Conservative    1;  Mismatches   14;  Indels    0;  Gaps    0;

Qy      10 VMNGFCRYPSHWRPLEQIRLLTKPERRLS 38
           ||| |    ||      ||||  || ||:
Db      21 VMNVFVSATSHQEDNVSIRLLVNPELRLT 49


RESULT   4
VST1_HEVME
ID   VST1_HEVME     STANDARD;      PRT;   123 AA.
AC   Q03499;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   01-JUN-1994 (Rel. 29, Last annotation update)
DE   STRUCTURAL PROTEIN 1.
OS   Hepatitis E virus (strain Mexico) (HEV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage.
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 6, 2002, 09:16:46 ; Search time 77.97 Seconds
(without alignments)
45.918 Million cell updates/sec

Title: US-09-897-042-16
Perfect score: 262
Sequence: 1 MNLVPMNPLVMNGFCRYPSH.....RLLTKPERRLSWLLPPLSNN 47

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	55.5	21.2	281	2	F75216	hypothetical prote
2	55	21.0	421	2	S36799	calreticulin precu
3	55	21.0	455	2	A53565	hydroxymethylgluta
4	54.5	20.8	279	2	D71453	hypothetical prote
5	54.5	20.8	362	2	A83967	carbamoyl-phosphat
6	54.5	20.8	553	2	C81024	hypothetical prote
7	54.5	20.8	558	2	F81967	probable inner mem
8	54	20.6	152	2	S38247	hypothetical prote
9	54	20.6	226	2	H64672	uroporphyrinogen I
10	54	20.6	970	2	S01352	type III site-spec
11	53.5	20.4	123	1	C44212	structural protein
12	53.5	20.4	256	2	T10821	photosynthetic rea
13	53	20.2	309	1	EUSMAG	agarase (EC 3.2.1.
14	53	20.2	911	1	HXADB3	hexon protein - bo
15	53	20.2	1215	2	T43916	chitinase A [impor
16	53	20.2	7829	2	T15789	hypothetical prote
17	52.5	20.0	282	1	NDHU1	deoxyribonuclease
18	52.5	20.0	452	2	JC4100	hydroxyindole O-me
19	52.5	20.0	4845	2	T31067	BIR repeat contain
20	52	19.8	400	2	S43376	calreticulin, brai
21	52	19.8	416	1	S06763	calreticulin precu
22	52	19.8	416	2	JH0819	calreticulin precu
23	52	19.8	417	1	A37047	calreticulin precu
24	52	19.8	418	1	A34154	calreticulin precu
25	52	19.8	837	2	H82970	hypothetical prote
26	52	19.8	1396	1	VCBE40	major capsid prote
27	52	19.8	1520	2	T30820	carbamoyl-phosphat
28	51.5	19.7	375	1	NRECD	ribonuclease III (
29	51.5	19.7	375	2	E85791	RNase D, processes
30	51.5	19.7	473	2	D83611	probable transcrip
31	51.5	19.7	805	2	S68441	leptin receptor, s
32	51.5	19.7	892	2	S68439	leptin receptor, s
33	51.5	19.7	894	2	S68437	leptin receptor, s
34	51.5	19.7	900	2	S68440	leptin receptor, s
35	51.5	19.7	953	2	S12894	pyruvate,orthophos
36	51.5	19.7	953	2	S56650	pyruvate,orthophos
37	51.5	19.7	1162	2	S68438	leptin receptor, s
38	51.5	19.7	1265	1	A37967	neural cell adhesi
39	51	19.5	278	2	S40849	l-rhamnose operon
40	51	19.5	278	2	G86079	positive regulator
41	51	19.5	280	2	T22425	hypothetical prote
42	51	19.5	339	2	E83496	hypothetical prote
43	51	19.5	388	2	T32307	hypothetical prote
44	51	19.5	411	2	S29129	calreticulin precu
45	51	19.5	539	2	T38454	methionine--tRNA l
46	51	19.5	711	2	F86373	protein T23E23.12
47	51	19.5	876	2	T51951	gamma-adaptin 1 [i
48	51	19.5	905	2	T00475	probable disease r
49	51	19.5	1186	2	T19050	hypothetical prote
50	50.5	19.3	92	2	C83288	hypothetical prote

ALIGNMENTS

RESULT 1
F75216
hypothetical protein PAB2181 - Pyrococcus abyssi (strain Orsay)
C;Species: Pyrococcus abyssi
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C;Accession: F75216
R;anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A;Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome s
A;Reference number: A75001
A;Accession: F75216
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-281 <KAW>
A;Cross-references: GB:AJ248283; GB:AL096836; NID:g5457433; PIDN:CAB49181.1; PID:g545
A;Experimental source: strain Orsay
C;Genetics:
A;Gene: PAB2181

Query Match 21.2%; Score 55.5; DB 2; Length 281;
Best Local Similarity 30.2%; Pred. No. 13;
Matches 13; Conservative 9; Mismatches 14; Indels 7; Gaps 2;

```
Qy       4 VPMNPLV------MNGFCRYPSHWRPLEQIRLLTKPERRLSWL 40
           ||:||:       :     : : :|| | | |:  |  ||:|:
Db       5 VPLNPITRSEIHQLESLLLFATLFRP-EVIELIKDPAERLTWV 46
```

RESULT 2
S36799
calreticulin precursor, brain isoform 2 - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 10-Dec-1993 #sequence_revision 23-Mar-1995 #text_change 13-Aug-1999
C;Accession: S36799; S36800
R;Liu, N.; Fine, R.E.; Johnson, R.J.
Biochim. Biophys. Acta 1202, 70-76, 1993
A;Title: Comparison of cDNAs from bovine brain coding for two isoforms of calreticuli
A;Reference number: S36799; MUID:93385184
A;Accession: S36799
A;Molecule type: mRNA
A;Residues: 1-421 <LIU>
A;Cross-references: GB:L13462; NID:g348693; PIDN:AAC37307.1; PID:g348694
A;Experimental source: brain, clone 9.4
A;Accession: S36800
A;Molecule type: protein
A;Residues: 35-45 <LI2>

```
C;Superfamily: calreticulin
C;Keywords: calcium binding; glycoprotein
F;1-34/Domain: signal sequence #status predicted <SIG>
F;35-421/Product: calreticulin, brain isoform 2 #status predicted <MAT>
F;418-421/Region: endoplasmic reticulum retention signal
F;141-167/Disulfide bonds: #status predicted
F;183/Binding site: carbohydrate (Asn) (covalent) #status predicted


  Query Match             21.0%;  Score 55;  DB 2;  Length 421;
  Best Local Similarity   30.8%;  Pred. No. 24;
  Matches   12;  Conservative    6;  Mismatches   15;  Indels    6;  Gaps    2;

Qy       8 PLVMNGFCRYPSHWRPLEQIRLLTKPERRLSWLLPPLSN 46
           ||: |    |   |:|    | :  || :  |: | : |
Db     268 PLIQNP--EYKGEWKP----RQIDNPEYKGIWIHPEIDN 300


RESULT   3
A53565
hydroxymethylglutaryl-CoA synthase (EC 4.1.3.5) - German cockroach
C;Species: Blattella germanica (German cockroach)
C;Date: 07-Oct-1994 #sequence_revision 07-Oct-1994 #text_change 21-Jul-2000
C;Accession: A53565
R;Buesa, C.; Martinez-Gonzalez, J.; Casals, N.; Haro, D.; Piulachs, M.D.; Belles, X.; He
J. Biol. Chem. 269, 11707-11713, 1994
A;Title: Blattella germanica has two HMG-CoA synthase genes. Both are regulated in the c
A;Reference number: A53565; MUID:94216267
A;Accession: A53565
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-455 <BUE>
A;Cross-references: GB:X77516; NID:g488133; PIDN:CAA54652.1; PID:g488134
C;Superfamily: hydroxymethylglutaryl-CoA synthase
C;Keywords: carbon-carbon lyase; coenzyme A; oxo-acid-lyase


  Query Match             21.0%;  Score 55;  DB 2;  Length 455;
  Best Local Similarity   45.5%;  Pred. No. 26;
  Matches   10;  Conservative    6;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       9 LVMNGFCRYPSHWRPLEQIRLL 30
           ||:| | :||  :: |:|:| |
Db     269 LVLNDFIQYPEKYQDLQQLRNL 290


RESULT   4
D71453
hypothetical protein PH0283 - Pyrococcus horikoshii
C;Species: Pyrococcus horikoshii
C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 21-Jul-2000
C;Accession: D71453
R;Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin
 M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi
DNA Res. 5, 55-76, 1998
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a
A;Reference number: A71000; MUID:98344137
A;Accession: D71453
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-279 <KAW>
A;Cross-references: GB:AP000001; NID:g3236128; PIDN:BAA29355.1; PID:g3256672
A;Experimental source: strain OT3
A;Note: this accession replaces an interim accession for a sequence replaced by GenBank
C;Genetics:
A;Gene: PH0283


  Query Match             20.8%;  Score 54.5;  DB 2;  Length 279;
  Best Local Similarity   27.9%;  Pred. No. 17;
  Matches   12;  Conservative   10;  Mismatches   14;  Indels    7;  Gaps    2;

Qy       4 VPMNPLV------MNGFCRYPSHWRPLEQIRLLTKPERRLSWL 40
           :|:||:       :     : : :|| | | |:  |  ||:|:
Db      11 IPLNPITRSEIHQLESLLLFATLFRP-EVIELIKDPAERLTWV 52


RESULT   5
A83967
carbamoyl-phosphate synthetase (glutaminase subunit) pyrAA [imported] - Bacillus halo
C;Species: Bacillus halodurans
C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 08-Dec-2000
C;Accession: A83967
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; H
Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a
A;Reference number: A83650; MUID:20263314
A;Accession: A83967
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-362 <STO>
A;Cross-references: GB:AP001515; GB:BA000004; NID:g10174886; PIDN:BAB06256.1; GSPDB:G
A;Experimental source: strain C-125
C;Genetics:
A;Gene: pyrAA
C;Superfamily: carbamoyl-phosphate synthase (glutamine-hydrolyzing) small chain; carb


  Query Match             20.8%;  Score 54.5;  DB 2;  Length 362;
  Best Local Similarity   34.1%;  Pred. No. 23;
  Matches   15;  Conservative    7;  Mismatches   13;  Indels    9;  Gaps    2;

Qy       6 MNP----LVMNGFCRYPSHWRPLEQIRLLTKPERRLSWLLPPLS 45
           |||    |::   .|  ||:||  | :  | | ::     :| ||
Db      71 MNPAIHGLIVKEACDIPSNWRSEESLDSLLKAKQ-----IPGLS 109


RESULT   6
C81024
hypothetical protein NMB1954 [imported] - Neisseria meningitidis (strain MC58 serogro
C;Species: Neisseria meningitidis
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C;Accession: C81024
R;Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen,
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A;Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.;
A;Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A;Reference number: A81000; MUID:20175755
A;Accession: C81024
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-553 <TET>
A;Cross-references: GB:AE002543; GB:AE002098; NID:g7227203; PIDN:AAF42283.1; PID:g722
A;Experimental source: serogroup B, strain MC58
C;Genetics:
A;Gene: NMB1954


  Query Match             20.8%;  Score 54.5;  DB 2;  Length 553;
  Best Local Similarity   30.9%;  Pred. No. 37;
  Matches   17;  Conservative    8;  Mismatches   17;  Indels   13;  Gaps    3;

Qy       4 VPMNPLVMNGFCR---YPSHWRPLEQIRLL-------TKPER---RLSWLLPPLS 45
           :|  || :   ||   : : |  |  : :|         |:|    | ||||||:
Db     281 LPALPLAVWTVCRTRLFSTDWGILGVVWMLAVLVLLAVNPQRFQDNLVWLLPPLA 335


RESULT   7
F81967
probable inner membrane protein NMA0497 [imported] - Neisseria meningitidis (strain Z
C;Species: Neisseria meningitidis
C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C;Accession: F81967
R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mo
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: January 6, 2002, 11:02:23 ; Search time 121.65 Seconds
(without alignments)
268.087 Million cell updates/sec

Title: US-09-897-042-15
Perfect score: 144
Sequence: 1 atgaacttggtaccgatgaa.....caccgctgagcaataactag 144

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	108.2	75.1	3984	1	US-08-040-753-1	Sequence 1, Appli
	2	73.8	51.2	8501	3	US-08-793-900-1	Sequence 1, Appli
	3	71.2	49.4	2220	2	US-08-864-224-1	Sequence 1, Appli
	4	69.2	48.1	278	1	US-08-178-477B-33	Sequence 33, Appl
	5	68.6	47.6	6727	2	US-08-125-462-2	Sequence 2, Appli
	6	68.6	47.6	6727	2	US-08-891-848-2	Sequence 2, Appli
	7	68.4	47.5	4819	1	US-08-450-257-20	Sequence 20, Appl
	8	68.4	47.5	4819	1	US-08-450-246-20	Sequence 20, Appl
	9	68.4	47.5	4819	1	US-08-450-098-20	Sequence 20, Appl
	10	68.4	47.5	4819	1	US-08-451-233-20	Sequence 20, Appl
	11	68.4	47.5	4819	1	US-08-450-236-20	Sequence 20, Appl
	12	68.4	47.5	5574	1	US-08-450-257-22	Sequence 22, Appl
	13	68.4	47.5	5574	1	US-08-450-246-22	Sequence 22, Appl
	14	68.4	47.5	5574	1	US-08-450-098-22	Sequence 22, Appl
	15	68.4	47.5	5574	1	US-08-451-233-22	Sequence 22, Appl
	16	68.4	47.5	5574	1	US-08-450-236-22	Sequence 22, Appl
	17	68.2	47.4	375	3	US-08-870-370-7	Sequence 7, Appli
	18	68.2	47.4	375	3	US-08-870-370-8	Sequence 8, Appli
	19	68.2	47.4	375	3	US-08-870-370-9	Sequence 9, Appli
	20	68	47.2	1010	4	US-09-070-408-1	Sequence 1, Appli
	21	68	47.2	2735	2	US-08-920-095-1	Sequence 1, Appli
	22	68	47.2	2735	5	PCT-US96-05800-1	Sequence 1, Appli
c	23	68	47.2	4411	2	US-08-929-967-5	Sequence 5, Appli
	24	68	47.2	5248	5	PCT-US96-05611A-21	Sequence 21, Appl
	25	68	47.2	5443	2	US-08-929-967-1	Sequence 1, Appli
	26	68	47.2	5616	2	US-08-929-967-3	Sequence 3, Appli
	27	68	47.2	6363	2	US-08-929-967-6	Sequence 6, Appli
c	28	68	47.2	6619	2	US-08-683-007A-3	Sequence 3, Appli
	29	67.8	47.1	6873	4	US-09-131-028A-1	Sequence 1, Appli
c	30	67.8	47.1	6873	4	US-09-131-028A-8	Sequence 8, Appli
c	31	67.8	47.1	8430	4	US-09-131-028A-6	Sequence 6, Appli
	32	67.8	47.1	8430	4	US-09-131-028A-10	Sequence 10, Appl
	33	67.4	46.8	7659	4	US-09-128-314-4	Sequence 4, Appli
	34	67.4	46.8	8157	4	US-09-128-314-3	Sequence 3, Appli
	35	67	46.5	111	2	US-08-606-288-9	Sequence 9, Appli
	36	67	46.5	111	3	US-09-347-483-9	Sequence 9, Appli
	37	67	46.5	234	2	US-08-687-865A-20	Sequence 20, Appl
	38	67	46.5	234	4	US-09-043-711-20	Sequence 20, Appl
	39	67	46.5	600	2	US-08-125-462-1	Sequence 1, Appli
	40	67	46.5	600	2	US-08-891-848-1	Sequence 1, Appli
	41	67	46.5	2029	4	US-09-136-574A-46	Sequence 46, Appl
	42	67	46.5	4410	1	US-08-594-469-1	Sequence 1, Appli
	43	67	46.5	4410	2	US-08-906-957-1	Sequence 1, Appli
	44	67	46.5	4977	1	US-08-450-257-14	Sequence 14, Appl
	45	67	46.5	4977	1	US-08-450-257-17	Sequence 17, Appl

ALIGNMENTS

RESULT 1
US-08-040-753-1
; Sequence 1, Application US/08040753
; Patent No. 5464745
; GENERAL INFORMATION:
; APPLICANT: Mierendorf, Robert
; APPLICANT: Garber, Richard
; APPLICANT: No. 5464745y, Robert
; APPLICANT: Hammer, Beth
; TITLE OF INVENTION: Protein Ligand Binding
; TITLE OF INVENTION: Region Mapping System
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles and Brady
; STREET: 1 South Pinckney St., Box 2113
; CITY: Madison
; STATE: WI
; COUNTRY: USA
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/040,753
; FILING DATE: 19930331
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 70-399-9001-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 608-251-5000
; TELEFAX: 608-251-9166
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3984 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; IMMEDIATE SOURCE:
; CLONE: pTOPE-1b(+)
; FEATURE:
; NAME/KEY: promoter
; LOCATION: 67..85

```
;      OTHER INFORMATION:  /function= "T7 Promoter"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  134..139
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  148..1092
;      OTHER INFORMATION:  /product= "5' end of T7 gene 10
;      OTHER INFORMATION:  fusion protein"/note= "This coding region is interrupted
;      OTHER INFORMATION:  during cloning by insertion of putative epitope encoding
;      OTHER INFORMATION:  DNA.  "
US-08-040-753-1


  Query Match            75.1%;  Score 108.2;  DB 1;  Length 3984;
  Best Local Similarity  97.3%;  Pred. No. 2.7e-28;
  Matches  110;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

 y      32 tgaacgaattctgcagatatccatcacactggcggccgctcgagcagatccggctgctaa 91
           ||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     980 TGCTGGAATTCTGCAGATATCCATCACACTGGCGGCCGCTCGAGCAGATCCGGCTGCTAA 1039

Qy      92 caaagcccgaaaggaagctgagttggctgctgccaccgctgagcaataactag 144
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1040 CAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAG 1092


RESULT   2
US-08-793-900-1
; Sequence 1, Application US/08793900
; Patent No. 6143518
;  GENERAL INFORMATION:
;    APPLICANT:  CAMERON, Beatrice
;    APPLICANT:  CROUZET, Joel
;    TITLE OF INVENTION:  METHOD FOR THE PRODUCTION OF RECOMBINANT
;    TITLE OF INVENTION:  PROTEINS, PLASMIDS AND MODIFIED CELLS
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Rhone-Poulenc Rorer Inc.
;      STREET:  500 Arcola Road, Mailstop 3C43
;      CITY:  Collegeville
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19426
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/793,900
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/FR95/01178
;      FILING DATE:  14-SEP-1995
;      APPLICATION NUMBER:  FR 94/11049
;      FILING DATE:  16-SEP-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Savitzky Esq., Martin F.
;      REGISTRATION NUMBER:  29,699
;      REFERENCE/DOCKET NUMBER:  ST94069-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (610) 454-3816
;      TELEFAX:  (610) 454-3808
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8501 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  circular
;    MOLECULE TYPE:  cDNA
US-08-793-900-1


  Query Match            51.2%;  Score 73.8;  DB 3;  Length 8501;
  Best Local Similarity  70.2%;  Pred. No. 3.5e-16;
  Matches   99;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy       4 aacttggtaccgatgaacccactagtaatgaacgaattctgcagatatccatcacactgg 63
           ||| ||| |       | | | | |  || ||| ||    |  || |  || |
Db     505 AACGCGGTCCTCGGACTCACTATGGCCAGAAAGCAATCTTGTTTCTCCCCCTGCCAGTCT 564

Qy      64 cggccgctcgagcagatccggctgctaacaaagcccgaaaggaagctgagttggctgctg 123
           |  | | |  | ||||||||||||||||||||||||||||||||||||||||||||||||
Db     565 CTTCTGATTAAAGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTG 624

Qy     124 ccaccgctgagcaataactag 144
           ||||||||||||||||||||
Db     625 CCACCGCTGAGCAATAACTAG 645


RESULT   3
US-08-864-224-1
; Sequence 1, Application US/08864224
; Patent No. 5851808
;  GENERAL INFORMATION:
;    APPLICANT:  Elledge, Stephen J.
;    APPLICANT:  Liu, Qinghua
;    TITLE OF INVENTION:  Rapid Subcloning Using Site-Specific
;    TITLE OF INVENTION:  Recombination
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Medlen & Carroll, LLP
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/864,224
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ingolia, Diane E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  BCM-02681
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2220 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  circular
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "DNA"
US-08-864-224-1


  Query Match            49.4%;  Score 71.2;  DB 2;  Length 2220;
  Best Local Similarity  78.7%;  Pred. No. 1.6e-15;
  Matches   85;  Conservative    0;  Mismatches   23;  Indels    0;  Gaps    0;

Qy      37 gaattctgcagatatccatcacactggcggccgctcgagcagatccggctgctaacaaag 96
           | |||  | ||| |          ||  ||    || || ||||||||||||||||||
Db     716 GGATTGGGAAGACAATAGCAGGCATGCTGGGGATTCTAGAAGATCCGGCTGCTAACAAAG 775

Qy      97 cccgaaaggaagctgagttggctgctgccaccgctgagcaataactag 144
           ||||||||||||||||||||||||||||||||||||||||||||||||
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: January 6, 2002, 11:06:50 ; Search time 256.95 Seconds
(without alignments)
480.462 Million cell updates/sec

Title: US-09-897-042-15
Perfect score: 144
Sequence: 1 atgaacttggtaccgatgaa.....caccgctgagcaataactag 144

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_1101:*
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	108.2	75.1	3984	16	AAT07310	Vector plasmid pTO
2	83.2	57.8	1104	20	AAZ20204	Mycobacterium tube
3	80.4	55.8	8031	21	AAC65950	Human lung cancer-
4	79.4	55.1	5502	22	AAF68849	Human lung tumour
5	79.2	55.0	8970	21	AAA57847	Nucleotide sequenc
6	79	54.9	3572	20	AAZ19454	M. tuberculosis an
7	79	54.9	3572	20	AAZ19242	M. tuberculosis re
8	79	54.9	7507	21	AAA53971	M. tuberculosis ge
9	79	54.9	7676	19	AAV64567	M. tuberculosis fu
10	79	54.9	7676	19	AAV55801	Mycobacterium anti
11	79	54.9	7676	20	AAZ20198	Mycobacterium tube
12	79	54.9	7676	20	AAZ19368	M. tuberculosis fu
13	79	54.9	7676	20	AAZ19156	M. tuberculosis fu
14	79	54.9	8217	22	AAS03795	M. tuberculosis ve
15	75	52.1	945	15	AAQ55302	gp55-Asn-Gly-Pro-h
16	73.8	51.2	8501	17	AAT12907	Novel expression p
17	73.8	51.2	8793	22	AAS03794	M. tuberculosis ve
18	71.8	49.9	5770	21	AAA15043	Nucleotide sequenc
19	71.4	49.6	5870	21	AAA15044	Nucleotide sequenc
20	71.2	49.4	2220	21	AAZ58064	Vector plasmid pUN
21	70.4	48.9	5216	20	AAX07548	pKS90 TIR coupling
22	70	48.6	1237	21	AAA53600	M. tuberculosis an
23	69.8	48.5	363	19	AAV44696	pTREP cassette in
24	69.8	48.5	5142	22	AAF29645	Plasmid pL2mTFF1v1
25	69.8	48.5	5230	20	AAX07549	pTREX1 mutant beta
26	69.8	48.5	5230	21	AAA15041	Nucleotide sequenc
27	69.8	48.5	5497	22	AAF29646	Plasmid pT1mTFF1.
28	69.8	48.5	5906	21	AAA15042	Nucleotide sequenc
29	68.6	47.6	6727	20	AAZ09858	pET-11d-ANG-E6 DNA
30	68.6	47.6	6727	20	AAV63616	Nucleic acid seque
31	68.4	47.5	116	22	AAF82146	Plasmid pET-3a nuc
32	68.4	47.5	4819	15	AAQ44360	Plasmid pTATdeltaC
33	68.4	47.5	4819	18	AAT89768	Plasmid pTAT-delta
34	68.4	47.5	4819	18	AAT88234	Plasmid pTAT-delta
35	68.4	47.5	4819	18	AAT84576	Plasmid pTATdelCys
36	68.4	47.5	4819	19	AAV56594	Plasmid pTATdeltaC
37	68.4	47.5	5231	20	AAX07550	pTREX1A mutant bet
38	68.4	47.5	5574	15	AAQ44361	Plasmid pTATdeltaC
39	68.4	47.5	5574	18	AAT89770	Plasmid pTAT-delta
40	68.4	47.5	5574	18	AAT88235	Plasmid pTAT-delta
41	68.4	47.5	5574	18	AAT84578	Plasmid pTATdelCys
42	68.4	47.5	5574	19	AAV56596	Plasmid pTATdeltaC
c 43	68.4	47.5	6499	21	AAA64015	Nucleotide sequenc
44	68.2	47.4	375	19	AAV54056	DNA fragment 2 of
45	68.2	47.4	375	19	AAV54057	DNA fragment 3 of

ALIGNMENTS

```
RESULT   1
AAT07310
ID   AAT07310 standard; DNA; 3984 BP.
XX
AC   AAT07310;
XX
DT   18-MAR-1996  (first entry)
XX
DE   Vector plasmid pTOPE-1b(+).
XX
KW   Plasmid pTOPE-1b(+); vector; ligand binding domain; epitope mapping;
KW   antigen; Escherichia coli; ds; cyclic.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   promoter        67..85
FT                   /*tag= a
FT                   /function= T7 promoter
FT   RBS             134..139
FT                   /*tag= b
FT   CDS             148..1092
FT                   /*tag= c
FT                   /product= 5' end of T7 gene 10 fusion protein
FT                   /note= "this coding region is interrupted during
FT                           cloning by insertion of putative epitope
FT                           encoding DNA"
XX
PN   US5464745-A.
XX
PD   07-NOV-1995.
XX
PF   31-MAR-1993;   93US-0040753.
XX
```

PR 31-MAR-1993; 93US-0040753.
XX
PA (NOVA-) NOVAGEN INC.
XX
PI Garber R, Hammer B, Mierendorf R, Novy R;
XX
DR WPI; 1995-392610/50.
DR P-PSDB; AAR87027.
XX
PT Mapping ligand binding domains, esp. epitope(s), of proteins - by
PT expressing peptide(s) encoded by random gene fragments and testing
PT for ligand binding
XX
PS Disclosure; Column 11-16; 12pp; English.
XX
CC Vector plasmid pTOPE-1b(+) (AAT07310) is a high-copy-number version
CC of pBR322-derivative plasmid pET-17xb. It is used in a method for
CC mapping the ligand binding domains (LBD), esp. antigenic regions, of
CC a protein. Putative LBD-encoding DNA is inserted into the vector
CC such that the LBD is expressed as a fusion to the T7 gene 10 leader
CC (AAR87027). The fusion protein accumulates as inclusion bodies in
CC Escherichia coli host cells and can be screened for its ability to
CC bind a ligand.
XX
SQ Sequence 3984 BP; 960 A; 1004 C; 1007 G; 1013 T; 0 other;

Query Match 75.1%; Score 108.2; DB 16; Length 3984;
Best Local Similarity 97.3%; Pred. No. 7.7e-27;
Matches 110; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 32 tgaacgaattctgcagatatccatcacactggcggccgctcgagcagatccggctgctaa 91
|| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 980 tgctggaattctgcagatatccatcacactggcggccgctcgagcagatccggctgctaa 1039

Qy 92 caaagcccgaaaggaagctgagttggctgctgccaccgctgagcaataactag 144
||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1040 caaagcccgaaaggaagctgagttggctgctgccaccgctgagcaataactag 1092

RESULT 2
AAZ20204
ID AAZ20204 standard; DNA; 1104 BP.
XX
AC AAZ20204;
XX
DT 17-JAN-2000 (first entry)
XX
DE Mycobacterium tuberculosis antigen fusion protein Mtb36f DNA.
XX
KW Tuberculosis; antigen; fusion protein; Mtb36f; ERD14; DPV; MTI;
KW diagnosis; therapy; vaccine; immunogen; ss.
XX
OS Mycobacterium tuberculosis.
XX
PN WO9951748-A2.
XX
PD 14-OCT-1999.
XX
PF 07-APR-1999; 99WO-US07717.
XX
PR 07-APR-1998; 98US-0056556.
PR 30-DEC-1998; 98US-0223040.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky YAW, Alderson M, Campos-Neto A;
XX
DR WPI; 1999-601610/51.
DR P-PSDB; AAY32069.
XX
PT [redacted]ion proteins useful for diagnosis, prevention and treatment of
PT [redacted]osis -
XX
PS Example; Fig 11A-B; 83pp; English.
XX
CC This DNA sequence includes a coding region for a recombinant
CC Mycobacterium tuberculosis tri-antigen fusion protein (see AAY32069),
CC termed Mtb36f, composed of the antigens ERD14, DPV and MTI. The
CC DNA is useful for the recombinant production of the fusion protein.
CC Coding sequences for the antigens were modified by PCR in order
CC to facilitate their fusion and subsequent expression of the fusion
CC protein, and were then ligated to obtain the present polynucleotide.
CC The invention provides fusion proteins (see AAY32059-71) containing
CC at least 2 M. tuberculosis antigens. The new fusion proteins and
CC polynucleotides encoding them are useful as vaccines for preventing
CC tuberculosis (claimed), for diagnosis (via in vitro assays or
CC intradermal skin tests for detection of anti-M. tuberculosis
CC antibodies), monitoring of disease progression, and treatment of
CC tuberculosis. They are more effective immunogens than mixtures of
CC the individual protein components.
XX
SQ Sequence 1104 BP; 216 A; 347 C; 333 G; 208 T; 0 other;

Query Match 57.8%; Score 83.2; DB 20; Length 1104;
Best Local Similarity 96.6%; Pred. No. 1.4e-18;
Matches 85; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 32 tgaacgaattctgcagatatccatcacactggcggccgctcgagcagatccggctgctaa 91
|| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1017 tgctggaattctgcagatatccatcacactggcggccgctcgagcagatccggctgctaa 1076

Qy 92 caaagcccgaaaggaagctgagttggct 119
|||||||||||||||||||||||||||||
Db 1077 caaagcccgaaaggaagctgagttggct 1104

RESULT 3
AAC65950
ID AAC65950 standard; DNA; 8031 BP.
XX
AC AAC65950;
XX
DT 21-FEB-2001 (first entry)
XX
DE Human lung cancer-associated DNA for L762P expression construct.
XX
KW Lung cancer; therapy; treatment; human; tumor; immunogenic; cytostatic;
KW vaccine; detection; ss.
XX
OS Homo sapiens.
XX
PN WO200061612-A2.
XX
PD 19-OCT-2000.
XX
PF 03-APR-2000; 2000WO-US08896.
XX
PR 02-APR-1999; 99US-0285479.
PR 17-DEC-1999; 99US-0466396.
PR 30-DEC-1999; 99US-0476496.
PR 10-JAN-2000; 2000US-0480884.
PR 22-FEB-2000; 2000US-0510376.
XX
PA (CORI-) CORIXA CORP.
XX
PI Wang T, Fan L;
XX
DR WPI; 2000-628399/60.
XX
PT Isolated polypeptide comprising an immunogenic portion of a lung tumor
PT protein is used for detecting and monitoring progression of lung cancer
PT in a patient -
XX
PS Claim 1a; Page 211-213; 261pp; English.

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: January 6, 2002, 10:33:27 ; Search time 2642.04 Seconds
(without alignments)
585.681 Million cell updates/sec

Title: US-09-897-042-15
Perfect score: 144
Sequence: 1 atgaacttggtaccgatgaa.....caccgctgagcaataactag 144

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_htc:*
10: gb_est1:*
11: gb_est2:*
12: gb_htc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	56	38.9	602	13	B06973	B06973 cSRL-8b6-t
c	2	51.6	35.8	301	11	BF286760	BF286760 EST451351
c	3	51.6	35.8	301	11	BF286761	BF286761 EST451352
c	4	49	34.0	594	10	BE439485	BE439485 HTM1-116F
	5	48.2	33.5	333	10	AU055647	AU055647 AU055647
c	6	48	33.3	574	10	BE439542	BE439542 HTM1-179F
c	7	48	33.3	748	10	BE439434	BE439434 HTM1-057F
c	8	47.2	32.8	245	11	BF286756	BF286756 EST451347
c	9	45.4	31.5	359	11	BF286758	BF286758 EST451349
c	10	44.8	31.1	302	11	BF286757	BF286757 EST451348
c	11	44.8	31.1	348	10	BE035760	BE035760 MO16E07 M
c	12	44.4	30.8	220	11	BF286794	BF286794 EST451385
c	13	44.4	30.8	327	11	BF286781	BF286781 EST451372
c	14	43.6	30.3	332	11	BF286778	BF286778 EST451369
c	15	43.6	30.3	332	11	BF286780	BF286780 EST451371
c	16	43.6	30.3	713	13	AZ574086	AZ574086 324PvF10
c	17	43.2	30.0	722	13	AZ570378	AZ570378 273PvF12
c	18	43.2	30.0	761	13	AZ572147	AZ572147 301PvG11
c	19	43	29.9	247	11	BF286712	BF286712 EST451303
c	20	43	29.9	247	11	BF286714	BF286714 EST451305
c	21	42.8	29.7	184	11	BF286797	BF286797 EST451388
	22	42.8	29.7	378	11	BF286772	BF286772 EST451363
c	23	42.6	29.6	234	10	AI110514	AI110514 394 neona
	24	42.6	29.6	734	10	AW231897	AW231897 2A30N6 Su
c	25	42.6	29.6	891	10	BE205638	BE205638 AOB64 Oni
	26	42.4	29.4	272	10	AU055688	AU055688 AU055688
c	27	42.4	29.4	287	11	BF286768	BF286768 EST451359
c	28	42.4	29.4	430	11	BF286771	BF286771 EST451362
c	29	42.4	29.4	479	10	AA545726	AA545726 HBMSF2E4-
	30	42.4	29.4	881	13	AF307808	AF307808 AF307808
c	31	42.2	29.3	141	11	BF286696	BF286696 EST451287
c	32	42.2	29.3	141	11	BF286749	BF286749 EST451340
c	33	42.2	29.3	300	11	BF286774	BF286774 EST451365
c	34	42.2	29.3	346	10	BE035761	BE035761 MO16E08 M
c	35	42.2	29.3	356	10	BE035762	BE035762 MO16E09 M
	36	42.2	29.3	361	10	AW697836	AW697836 EST0019 c
c	37	42.2	29.3	565	11	BI273540	BI273540 F19A1 Fee
	38	42.2	29.3	570	11	BG577408	BG577408 N120 SSH-
c	39	42.2	29.3	624	10	AJ410657	AJ410657 AJ410657
c	40	42	29.2	130	11	BG673766	BG673766 906 cilia
	41	42	29.2	183	10	AU055645	AU055645 AU055645
	42	42	29.2	206	10	AA389640	AA389640 HFBEST- H
c	43	42	29.2	221	10	AA389641	AA389641 HBCEST- H
c	44	42	29.2	247	10	AW608990	AW608990 PM3-PT004
	45	42	29.2	257	13	AF034178	AF034178 AF034178

ALIGNMENTS

RESULT 1
B06973
LOCUS B06973 602 bp DNA GSS 13-JUL-1996
DEFINITION cSRL-8b6-t cSRL flow sorted Chromosome 11 specific cosmid Homo sapiens genomic clone cSRL-8b6, DNA sequence.
ACCESSION B06973
VERSION B06973.1 GI:1416251
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 602)
AUTHORS Evans,G.A., Burbee,D., Davies,C., Hahner,L., Oliver,T., Gilbert,M., Jones,D., Ward,T., Gillilan,E., Schagemann,J., Probst,S. ., Harris ,J., DeFord,J., McFarland,J., Burzinski,K., Khan,M., Kupfer,K. and Garner,H.R.
TITLE Genomic Sequence Sampled Map of Chromosome 11
JOURNAL Unpublished (1996)
COMMENT Contact: Evans GA, Shane Probst
McDermott Center for Human Growth and Development
University of Texas Southwestern Medical Center At Dallas
5323 Harry Hines Blvd, Dallas TX 75235-8591
Tel: 214-648-1600
Fax: 214-648-1666
Email: gevans@utsw.swmed.edu, shane@mcdermott.swmed.edu
PCR PRimers
FORWARD: GAAAGGAGGAACTATATCCG
BACKWARD: ATTGGTTAAAAAATGAGCTG
Seq primer: T3
Class: cosmid ends
High quality sequence stop: 602.
FEATURES Location/Qualifiers
source 1. .602
/organism="Homo sapiens"

```
                     /db_xref="taxon:9606"
                     /clone="cSRL-8b6"
                     /clone_lib="cSRL flow sorted Chromosome 11 specific
                     cosmid"
                     /sex="female"
                     /cell_type="chimeric hamster somatic cell hybrid"
                     /note="Vector: sCos-1; Human Chromosome 11 specific cosmid
                     library prepared from flow sorted human  Chromosome 11
                     derived from Chinese Hampster Ovary (CHO) monochromosomal
                     somatic cell hybrid, J1"
BASE COUNT      148 a    126 c    151 g    143 t     34 others
ORIGIN


  Query Match             38.9%;  Score 56;  DB 13;  Length 602;
  Best Local Similarity   98.5%;  Pred. No. 1.7e-06;
  Matches   67;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy     78 gatccggctgct-aacaaagcccgaaaggaagctgagttggctgctgccaccgctgagca 136
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db    101 GATCCGGCTGCTNAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCA 160

Qy    137 ataactag 144
          ||||||||
Db    161 ATAACTAG 168


RESULT   2
BF286760/c
LOCUS       BF286760      301 bp    mRNA            EST       28-NOV-2000
DEFINITION  EST451351 Rat Gene Index, normalized rat, Rattus norvegicus cDNA
            Rattus norvegicus cDNA clone RGIFS57 3' sequence, mRNA sequence.
ACCESSION   BF286760
VERSION     BF286760.1  GI:11217830
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 301)
  AUTHORS   Malek,R.L., Cho,J., Lee,Y., Karamycheva,S., Parvizi,B., Pertea,G.,
            Sultana,R., Tsai,J., White,J., Quackenbush,J. and Lee,N.H.
  TITLE     Generation of ESTs from Normalized Rat Embryo, Bento Soares
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Lee, NH
            The Institute for Genomic Research
            9712, Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208
            Email: nhlee@tigr.org
            This clone is available through the ATCC, contact the ATCC
            tel#703-365-2700 for further information.
FEATURES             Location/Qualifiers
     source          1. .301
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /clone="RGIFS57"
                     /clone_lib="Rat Gene Index, normalized rat, Rattus
                     norvegicus cDNA"
                     /tissue_type="mixed tissue"
                     /lab_host="DH5-alpha"
                     /note="Vector: pT3T7Pac; Site_1: EcoR1; Site_2: Not1;
                     Combination of ROV, RBR, RKI, RLI, RPL, RLU, REM, RMU, RSP
                     , RHE, RPC, RPN"
BASE COUNT       72 a     60 c     72 g     97 t
ORIGIN


  Query Match             35.8%;  Score 51.6;  DB 11;  Length 301;
  Best Local Similarity   67.9%;  Pred. No. 3e-05;
  Matches   72;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy     32 tgaacgaattctgcagatatccatcacactggcggccgctcgagcagatccggctgctaa 91
          ||   ||||||||||||||||||||||||||||||||||||||||||  |||   |  | |
Db    213 TGCTGGAATTCTGCAGATATCCATCACACTGGCGGCCGCTCGAGCATGTCCAAGTATTGA 154

Qy     92 caaagcccgaaaggaagctgagttggctgctgccaccgctgagcaa 137
          ||| ||  | | | |   ||   | |||    |  |  |||| ||
Db    153 AAAAACCGCACATGGACAAGAACAGACTGTCAACTACCATGAGGAA 108


RESULT   3
BF286761/c
LOCUS       BF286761      301 bp    mRNA            EST       28-NOV-2000
DEFINITION  EST451352 Rat Gene Index, normalized rat, Rattus norvegicus cDNA
            Rattus norvegicus cDNA clone RGIFS58 3' sequence, mRNA sequence.
ACCESSION   BF286761
VERSION     BF286761.1  GI:11217831
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 301)
  AUTHORS   Malek,R.L., Cho,J., Lee,Y., Karamycheva,S., Parvizi,B., Pertea,G.,
            Sultana,R., Tsai,J., White,J., Quackenbush,J. and Lee,N.H.
  TITLE     Generation of ESTs from Normalized Rat Embryo, Bento Soares
  JOURNAL   Unpublished (2000)
COMMENT     Other_ESTs: EST451353
            Contact: Lee, NH
            The Institute for Genomic Research
            9712, Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208
            Email: nhlee@tigr.org
            This clone is available through the ATCC, contact the ATCC
            tel#703-365-2700 for further information.
FEATURES             Location/Qualifiers
     source          1. .301
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /clone="RGIFS58"
                     /clone_lib="Rat Gene Index, normalized rat, Rattus
                     norvegicus cDNA"
                     /tissue_type="mixed tissue"
                     /lab_host="DH5-alpha"
                     /note="Vector: pT3T7Pac; Site_1: EcoR1; Site_2: Not1;
                     Combination of ROV, RBR, RKI, RLI, RPL, RLU, REM, RMU, RSP
                     , RHE, RPC, RPN"
BASE COUNT       72 a     60 c     72 g     97 t
ORIGIN


  Query Match             35.8%;  Score 51.6;  DB 11;  Length 301;
  Best Local Similarity   67.9%;  Pred. No. 3e-05;
  Matches   72;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy     32 tgaacgaattctgcagatatccatcacactggcggccgctcgagcagatccggctgctaa 91
          ||   ||||||||||||||||||||||||||||||||||||||||||  |||   |  | |
Db    213 TGCTGGAATTCTGCAGATATCCATCACACTGGCGGCCGCTCGAGCATGTCCAAGTATTGA 154

Qy     92 caaagcccgaaaggaagctgagttggctgctgccaccgctgagcaa 137
          ||| ||  | | | |   ||   | |||    |  |  |||| ||
Db    153 AAAAACCGCACATGGACAAGAACAGACTGTCAACTACCATGAGGAA 108


RESULT   4
BE439485/c
LOCUS       BE439485      594 bp    mRNA            EST       25-JUL-2000
DEFINITION  HTM1-116F1 HTM1 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE439485
VERSION     BE439485.1  GI:9438967
KEYWORDS    EST.
SOURCE      human.
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: January 6, 2002, 11:00:34 ; Search time 1586.23 Seconds
(without alignments)
1497.635 Million cell updates/sec

Title: US-09-897-042-15
Perfect score: 144
Sequence: 1 atgaacttggtaccgatgaa.....caccgctgagcaataactag 144

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vi:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rod:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rod:*
36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

%

Result No.	Score	Query Match	Length	DB	ID	Description
1	108.2	75.1	3984	6	I15353	I15353 Sequence 1
2	79.4	55.1	5502	6	AX063158	AX063158 Sequence
3	75	52.1	717	6	A36847	A36847 Sequence 1
4	75	52.1	945	6	A36849	A36849 Sequence 3
5	73.8	51.2	8501	6	A49428	A49428 Sequence 1
6	71.2	49.4	2039	12	AF156991	AF156991 Cloning v
7	71.2	49.4	2060	12	AF156992	AF156992 Cloning v
8	71.2	49.4	2075	12	AF156990	AF156990 Cloning v
9	71.2	49.4	2220	6	AR067785	AR067785 Sequence
10	71.2	49.4	2220	12	AF143506	AF143506 Cloning v
11	71.2	49.4	2326	12	AF143508	AF143508 Cloning v
12	71.2	49.4	2490	12	AF143507	AF143507 Cloning v
13	71.2	49.4	2526	12	AF149259	AF149259 Cloning v
14	71.2	49.4	2547	12	AF149260	AF149260 Cloning v
15	71.2	49.4	2562	12	AF149258	AF149258 Cloning v
16	71.2	49.4	2669	12	AF149261	AF149261 Cloning v
17	71.2	49.4	2788	12	AF149263	AF149263 Cloning v
18	71.2	49.4	2811	12	AF149262	AF149262 Cloning v
19	71.2	49.4	2966	12	AF149266	AF149266 Cloning v
20	71.2	49.4	3028	12	AF149267	AF149267 Cloning v
21	71.2	49.4	3378	12	AF149265	AF149265 Cloning v
22	71.2	49.4	3417	12	AF149264	AF149264 Cloning v
23	70.4	48.9	5217	6	AX001271	AX001271 Sequence
24	70.2	48.8	3123	12	PMBC2TD	X77750 Plasmid pMB
25	70	48.6	3584	1	ECOLACAB	M74750 E.coli LacZ
26	70	48.6	3741	12	PKSM711	U04894 Cloning vec
27	70	48.6	5025	12	ECO7659	AJ007659 Cloning v
28	70	48.6	5041	12	PKSM713	U04895 Cloning vec
29	70	48.6	5041	12	PKSM715	U04896 Cloning vec
30	70	48.6	5071	12	PKSM710	U04893 Cloning vec
31	70	48.6	5406	12	ECO7660	AJ007660 Cloning v
32	69.8	48.5	383	6	A61284	A61284 Sequence 6
33	69.8	48.5	5142	6	AX069289	AX069289 Sequence
34	69.8	48.5	5230	6	AX001272	AX001272 Sequence
35	69.8	48.5	5505	6	AX069290	AX069290 Sequence
36	69.4	48.2	2863	12	AF013597	AF013597 Phagemid
37	69.2	48.1	278	6	AR009724	AR009724 Sequence
38	69	47.9	2828	12	CVPRSET6D	X54209 Cloning vec
39	69	47.9	2829	12	CVPRSET6C	X54208 Cloning vec
40	69	47.9	2830	12	CVPRSET6B	X54207 Cloning vec
41	69	47.9	2831	12	CVPRSET6A	X54206 Cloning vec
42	68.6	47.6	2730	1	PET81F1P	X62499 E.coli plas
43	68.6	47.6	2731	1	PET31F1P	X62498 E.coli plas
44	68.6	47.6	2891	12	CVT7NCO	U25273 Expression
45	68.6	47.6	6101	12	CVPSIT	U17139 Phagemid cl

ALIGNMENTS

RESULT 1
I15353
LOCUS I15353 3984 bp DNA PAT 02-APR-1996
DEFINITION Sequence 1 from patent US 5464745.
ACCESSION I15353
VERSION I15353.1 GI:1250261
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 3984)
AUTHORS Mierendorf,R., Garber,R., Novy,R. and Hammer,B.
TITLE Protein ligand binding region mapping system
JOURNAL Patent: US 5464745-A 1 07-NOV-1995;
FEATURES Location/Qualifiers
source 1. .3984
/organism="unknown"
BASE COUNT 960 a 1004 c 1007 g 1013 t
ORIGIN

```
  Query Match             75.1%;  Score 108.2;  DB 6;  Length 3984;
  Best Local Similarity   97.3%;  Pred. No. 9.7e-24;
  Matches  110;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      32 tgaacgaattctgcagatatccatcacactggcggccgctcgagcagatccggctgctaa 91
           ||   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     980 TGCTGGAATTCTGCAGATATCCATCACACTGGCGGCCGCTCGAGCAGATCCGGCTGCTAA 1039

Qy      92 caaagcccgaaaggaagctgagttggctgctgccaccgctgagcaataactag 144
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1040 CAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAG 1092


RESULT   2
AX063158
LOCUS       AX063158     5502 bp    DNA             PAT       24-JAN-2001
DEFINITION  Sequence 785 from Patent WO0100828.
ACCESSION   AX063158
VERSION     AX063158.1  GI:12541015
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5502)
  AUTHORS   Wang,T., Bangur,C.S., Lodes,M.J., Fanger,G.R., Vedvick,T.S.,
            Carter,D., Retter,M.W. and Mannion,J.
  TITLE     Compositions and methods for the therapy and diagnosis of lung
            cancer
  JOURNAL   Patent: WO 0100828-A 785 04-JAN-2001;
            CORIXA CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .5502
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT     1306 a   1479 c   1409 g   1308 t
ORIGIN


  Query Match             55.1%;  Score 79.4;  DB 6;  Length 5502;
  Best Local Similarity   84.5%;  Pred. No. 1.3e-14;
  Matches  109;  Conservative    0;  Mismatches    1;  Indels   19;  Gaps    1;

Qy      35 acgaattctgcagatatccatcacactggcggccgctcgagc---------------------- 76
           | ||||||||||||||||||||||||||||||||||||||
Db    5305 ATGAATTCTGCAGATATCCATCACACTGGCGGCCGCTCGAGCACCACCACCACCACCACT 5364

Qy      77 -agatccggctgctaacaaagcccgaaaggaagctgagttggctgctgccaccgctgagc 135
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5365 GAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGC 5424

Qy     136 aataactag 144
           |||||||||
Db    5425 AATAACTAG 5433


RESULT   3
A36847
LOCUS       A36847       717 bp    DNA             PAT       05-MAR-1997
DEFINITION  Sequence 1 from Patent WO9402510.
ACCESSION   A36847
VERSION     A36847.1  GI:2294105
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 717)
  AUTHORS   Bauer,W., Breckenridge,R., Cardinaux,F., Gombert,F., Gram,H.,
            Ramage,P., Schneider,H., Waelchli,R., Albert, Rainer and Lewis,I.
  TITLE     ANALOGS OF PTH
  JOURNAL   Patent: WO 9402510-A 1 03-FEB-1994;
            SANDOZ AG (AT)
COMMENT     Other publication HU 70459 951030
            Other publication CZ 9500088 951018
            Other publication AU 4156693 940120
            Other publication SK 4395 950607
            Other publication NZ 248137 951221
            Other publication GB 2269176 940202
            Other publication JP 6184198 940705
            Other publication CA 2100423 940116
            Other publication CN 1099801 950308
            Other publication DE 4393381T 950427
            Other publication NO 950123 950315
            Other publication ZA 9305126 950116
            Other publication FI 950171 950313.
FEATURES             Location/Qualifiers
     source          1. .717
                     /organism="unidentified"
                     /db_xref="taxon:32644"
     CDS             101. .562
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAA02267.1"
                     /db_xref="GI:2294106"
                     /translation="MSETKPKYNYVNNKELLQAIIDWKTELANNKDPNKVVRQNDTIG
                     LAIMLIAEGLSKRFNFSGYTQSWKQEMIADGIEASIKGLHNFDETKYKNPHAYITQAC
                     FNAFVQRGSIDPPSVSEIQLMHNLGKHLNSMERVEWLRKKLQDVHNFVALG"
BASE COUNT      245 a    137 c    144 g    191 t
ORIGIN


  Query Match             52.1%;  Score 75;  DB 6;  Length 717;
  Best Local Similarity   100.0%;  Pred. No. 3.2e-13;
  Matches   75;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      70 ctcgagcagatccggctgctaacaaagcccgaaaggaagctgagttggctgctgccaccg 129
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    563 CTCGAGCAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCG 622

Qy     130 ctgagcaataactag 144
           ||||||||||||||
Db    623 CTGAGCAATAACTAG 637


RESULT   4
A36849
LOCUS       A36849       945 bp    DNA             PAT       05-MAR-1997
DEFINITION  Sequence 3 from Patent WO9402510.
ACCESSION   A36849
VERSION     A36849.1  GI:2294107
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 945)
  AUTHORS   Bauer,W., Breckenridge,R., Cardinaux,F., Gombert,F., Gram,H.,
            Ramage,P., Schneider,H., Waelchli,R., Albert, Rainer and Lewis,I.
  TITLE     ANALOGS OF PTH
  JOURNAL   Patent: WO 9402510-A 3 03-FEB-1994;
            SANDOZ AG (AT)
COMMENT     Other publication HU 70459 951030
            Other publication CZ 9500088 951018
            Other publication AU 4156693 940120
            Other publication SK 4395 950607
            Other publication NZ 248137 951221
            Other publication GB 2269176 940202
            Other publication JP 6184198 940705
            Other publication CA 2100423 940116
            Other publication CN 1099801 950308
            Other publication DE 4393381T 950427
            Other publication NO 950123 950315
            Other publication ZA 9305126 950116
            Other publication FI 950171 950313.
FEATURES             Location/Qualifiers
     source          1. .945
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 6, 2002, 09:13:46 ; Search time 71.9 Seconds
(without alignments)
2.191 Million cell updates/sec

Title: US-09-897-042-14
Perfect score: 37
Sequence: 1 MNLVPMN 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	31	83.8	2071	4	US-09-415-522-6	Sequence 6, Appli
2	29	78.4	237	3	US-08-924-747-28	Sequence 28, Appl
3	29	78.4	237	4	US-09-247-373B-28	Sequence 28, Appl
4	29	78.4	237	4	US-09-296-715-28	Sequence 28, Appl
5	28	75.7	75	1	US-08-428-415-14	Sequence 14, Appl
6	28	75.7	75	1	US-08-379-685-14	Sequence 14, Appl
7	28	75.7	75	2	US-08-854-029-14	Sequence 14, Appl
8	28	75.7	75	4	US-08-428-762-14	Sequence 14, Appl
9	28	75.7	211	1	US-08-428-415-9	Sequence 9, Appli
10	28	75.7	211	1	US-08-379-685-9	Sequence 9, Appli
11	28	75.7	211	2	US-08-854-029-9	Sequence 9, Appli
12	28	75.7	211	4	US-08-428-762-9	Sequence 9, Appli
13	28	75.7	396	3	US-09-082-310-2	Sequence 2, Appli
14	28	75.7	2500	2	US-08-801-263A-2	Sequence 2, Appli
15	28	75.7	2500	3	US-09-102-248-2	Sequence 2, Appli
16	28	75.7	2512	2	US-08-801-263A-9	Sequence 9, Appli
17	28	75.7	2512	3	US-09-102-248-9	Sequence 9, Appli
18	28	75.7	2517	2	US-08-801-263A-5	Sequence 5, Appli
19	28	75.7	2517	3	US-09-102-248-5	Sequence 5, Appli
20	27	73.0	214	4	US-09-587-066-6	Sequence 6, Appli
21	27	73.0	273	1	US-08-118-270-63	Sequence 63, Appl
22	27	73.0	273	5	PCT-US93-08528-63	Sequence 63, Appl
23	27	73.0	398	1	US-08-446-777-2	Sequence 2, Appli
24	27	73.0	671	6	5266464-2	Patent No. 5266464
25	27	73.0	694	2	US-08-701-240-2	Sequence 2, Appli
26	27	73.0	694	4	US-09-138-236-2	Sequence 2, Appli
27	27	73.0	695	2	US-08-701-240-4	Sequence 4, Appli
28	27	73.0	695	4	US-09-138-236-4	Sequence 4, Appli
29	27	73.0	767	1	US-08-446-777-6	Sequence 6, Appli
30	27	73.0	767	1	US-08-446-777-8	Sequence 8, Appli
31	27	73.0	1012	1	US-08-219-262B-10	Sequence 10, Appl
32	27	73.0	1012	3	US-09-031-655-10	Sequence 10, Appl
33	27	73.0	1013	2	US-08-708-541A-34	Sequence 34, Appl
34	27	73.0	1824	2	US-08-680-327-3	Sequence 3, Appli
35	27	73.0	1824	4	US-09-228-246-2	Sequence 2, Appli
36	26	70.3	9	3	US-09-075-257A-1	Sequence 1, Appli
37	26	70.3	9	4	US-09-534-639-1	Sequence 1, Appli
38	26	70.3	10	3	US-09-075-257A-10	Sequence 10, Appl
39	26	70.3	10	4	US-09-534-639-10	Sequence 10, Appl
40	26	70.3	15	4	US-08-256-104-5	Sequence 5, Appli
41	26	70.3	15	4	US-08-408-930A-4	Sequence 4, Appli
42	26	70.3	25	3	US-09-075-257A-17	Sequence 17, Appl
43	26	70.3	25	4	US-09-534-639-17	Sequence 17, Appl
44	26	70.3	28	3	US-09-075-257A-15	Sequence 15, Appl
45	26	70.3	28	3	US-09-075-257A-16	Sequence 16, Appl
46	26	70.3	28	4	US-09-534-639-15	Sequence 15, Appl
47	26	70.3	28	4	US-09-534-639-16	Sequence 16, Appl
48	26	70.3	29	3	US-09-075-257A-13	Sequence 13, Appl
49	26	70.3	29	3	US-09-075-257A-14	Sequence 14, Appl
50	26	70.3	29	3	US-09-075-257A-18	Sequence 18, Appl

ALIGNMENTS

RESULT 1
US-09-415-522-6
; Sequence 6, Application US/09415522A
; Patent No. 6291660
; GENERAL INFORMATION:
; APPLICANT: Gaffney, Thomas
; APPLICANT: Wendland, Juergen
; APPLICANT: Philippsen, Peter
; TITLE OF INVENTION: No. 6291660el Fungal Genes Required For No. 6291660mal Growth
; TITLE OF INVENTION: Development
; FILE REFERENCE: CGC2046
; CURRENT APPLICATION NUMBER: US/09/415,522A
; CURRENT FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 2071
; TYPE: PRT
; ORGANISM: Ashbya gossypii
US-09-415-522-6

Query Match 83.8%; Score 31; DB 4; Length 2071;
Best Local Similarity 85.7%; Pred. No. 1.9e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MNLVPMN 7
||| |||
Db 718 MNLPPMN 724

RESULT 2
US-08-924-747-28
; Sequence 28, Application US/08924747
; Patent No. 6063570
; GENERAL INFORMATION:
; APPLICANT: MCGONIGLE, BRIAN
; APPLICANT: O'KEEFE, DANIEL
; TITLE OF INVENTION: SOYBEAN GLUTATHIONE-S-TRANSFERASE
; TITLE OF INVENTION: ENZYMES
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E.I. DU PONT DE NEMOURS AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON

```
;       STATE:  DELAWARE
;       COUNTRY:  UNITED STATES OF AMERICA
;       ZIP:  19898
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  DISKETTE, 3.50 INCH
;       COMPUTER:  IBM PC COMPATIBLE
;       OPERATING SYSTEM:  MICROSOFT WORD FOR WINDOWS 95
;       SOFTWARE:  MICROSOFT WORD VERSION 7.0A
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/924,747
;       FILING DATE:
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  FLOYD, LINDA AXAMETHY
;       REGISTRATION NUMBER:  33,692
;       REFERENCE/DOCKET NUMBER:  CL-1108
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  302-892-8112
;       TELEFAX:  302-773-0164
;   INFORMATION FOR SEQ ID NO:  28:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  237 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  not relevant
;       TOPOLOGY:  not relevant
;     MOLECULE TYPE:  protein
;     ORIGINAL SOURCE:
;       TISSUE TYPE:  SOYBEAN
;     IMMEDIATE SOURCE:
;       CLONE:  SE1.PK0017.F5
US-08-924-747-28


  Query Match             78.4%;  Score 29;  DB 3;  Length 237;
  Best Local Similarity   71.4%;  Pred. No. 50;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MNLVPMN 7
           :||||:|
Db      55 INLVPIN 61


RESULT   3
US-09-247-373B-28
; Sequence 28, Application US/09247373B
; Patent No. 6168954
; GENERAL INFORMATION:
;  APPLICANT:  MCGONIGLE, BRIAN
;  APPLICANT:  O'KEEFE, DANIEL
;  TITLE OF INVENTION:  SOYBEAN GLUTATHIONE-S-TRANSFERASE ENZYMES
;  FILE REFERENCE:  CL-1108-A
;  CURRENT APPLICATION NUMBER: US/09/247,373B
;  CURRENT FILING DATE:  1999-02-10
;  PRIOR APPLICATION NUMBER:  08/924,747
;  PRIOR FILING DATE:  1997-09-05
;  NUMBER OF SEQ ID NOS:  56
;  SOFTWARE:  Microsoft Office 97
; SEQ ID NO 28
;   LENGTH: 237
;   TYPE: PRT
;   ORGANISM: SOYBEAN
US-09-247-373B-28


  Query Match             78.4%;  Score 29;  DB 4;  Length 237;
  Best Local Similarity   71.4%;  Pred. No. 50;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MNLVPMN 7
           :||||:|
Db      55 INLVPIN 61


RESULT   4
US-09-296-715-28
; Sequence 28, Application US/09296715
; Patent No. 6171839
;  GENERAL INFORMATION:
;    APPLICANT:  MCGONIGLE, BRIAN
;    APPLICANT:  O'KEEFE, DANIEL
;    TITLE OF INVENTION:  SOYBEAN GLUTATHIONE-S-TRANSFERASE
;    TITLE OF INVENTION:  ENZYMES
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  E.I. DU PONT DE NEMOURS AND COMPANY
;      STREET:  1007 MARKET STREET
;      CITY:  WILMINGTON
;      STATE:  DELAWARE
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  19898
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  DISKETTE, 3.50 INCH
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  MICROSOFT WORD FOR WINDOWS 95
;      SOFTWARE:  MICROSOFT WORD VERSION 7.0A
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/296,715
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  FLOYD, LINDA AXAMETHY
;      REGISTRATION NUMBER:  33,692
;      REFERENCE/DOCKET NUMBER:  CL-1108
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  302-892-8112
;      TELEFAX:  302-773-0164
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  237 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  protein
;    ORIGINAL SOURCE:
;      TISSUE TYPE:  SOYBEAN
;    IMMEDIATE SOURCE:
;      CLONE:  SE1.PK0017.F5
US-09-296-715-28


  Query Match             78.4%;  Score 29;  DB 4;  Length 237;
  Best Local Similarity   71.4%;  Pred. No. 50;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MNLVPMN 7
           :||||:|
Db      55 INLVPIN 61


RESULT   5
US-08-428-415-14
; Sequence 14, Application US/08428415
; Patent No. 5756335
;  GENERAL INFORMATION:
;    APPLICANT:  Cold Spring Harbor Laboratory
;    TITLE OF INVENTION:  No. 5756335el Human cdc25 Genes, Encoded
;    TITLE OF INVENTION:  Products and Uses Thereof
;    NUMBER OF SEQUENCES:  31
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  60 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02109
;    COMPUTER READABLE FORM:
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 6, 2002, 09:13:46 ; Search time 103.34 Seconds
(without alignments)
5.018 Million cell updates/sec

Title: US-09-897-042-14
Perfect score: 37
Sequence: 1 MNLVPMN 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : A_Geneseq_1101:*
1: /SIDS2/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	31	83.8	2071	21	AAY84686	Amino acid sequenc
2	29	78.4	237	21	AAB07839	Amino acid sequenc
3	29	78.4	237	21	AAB03744	Clone sel.pk0017.f
4	29	78.4	237	21	AAY79525	Soybean glutathion
5	29	78.4	237	22	AAB66743	Soybean type IV GS
6	29	78.4	366	22	AAG91960	C glutamicum prote
7	29	78.4	424	21	AAG20275	Arabidopsis thalia
8	29	78.4	424	21	AAG42884	Arabidopsis thalia
9	29	78.4	436	21	AAG42883	Arabidopsis thalia
10	29	78.4	437	21	AAG20274	Arabidopsis thalia
11	29	78.4	933	22	AAB50437	Arabidopsis thalia
12	28	75.7	48	20	AAY14446	Human secreted pro
13	28	75.7	101	18	AAW16679	Lawsonia intracell
14	28	75.7	215	21	AAB16004	E. coli proliferat
15	28	75.7	344	22	AAM39369	Human polypeptide
16	28	75.7	353	22	AAM41155	Human polypeptide
17	28	75.7	364	22	AAG82991	S. epidermidis ope
18	28	75.7	391	21	AAB43612	Human cancer assoc
19	28	75.7	391	22	AAB94155	Human protein sequ
20	28	75.7	396	21	AAY44277	Human nucleic acid
21	28	75.7	396	22	AAM38840	Human polypeptide
22	28	75.7	407	22	AAM40626	Human polypeptide
23	28	75.7	807	19	AAW70461	South African Arbo
24	28	75.7	807	19	AAW70463	Girdwood S.A.virus
25	28	75.7	807	19	AAW70465	Sindbis virus nsP2
26	27	73.0	48	22	AAM33402	Peptide #7439 enco
27	27	73.0	56	22	AAB59287	GlmU protein. Uni
28	27	73.0	110	22	AAE05970	Protein kinase C (
29	27	73.0	115	22	AAE05975	Rat protein kinase
30	27	73.0	128	22	AAE05972	Rat protein kinase
31	27	73.0	136	21	AAB01483	Lectin polypeptide
32	27	73.0	170	21	AAY96467	Partial Veronia 4-
33	27	73.0	214	22	AAE00712	Rhodopseudomonas s
34	27	73.0	216	22	AAG98603	Mouse olfactory re
35	27	73.0	216	22	AAG98606	Mouse olfactory re
36	27	73.0	248	22	AAB87636	Bovine mammary tis
37	27	73.0	273	15	AAR48742	G-protein coupled
38	27	73.0	273	17	AAW02714	G-protein coupled
39	27	73.0	275	22	AAB95047	Human protein sequ
40	27	73.0	278	21	AAG36698	Arabidopsis thalia
41	27	73.0	278	21	AAG37341	Arabidopsis thalia
42	27	73.0	278	21	AAG44322	Arabidopsis thalia
43	27	73.0	279	21	AAB42870	Human ORFX ORF2634
44	27	73.0	317	13	AAR27870	Odorant receptor c
45	27	73.0	317	22	AAG73007	Olfactory receptor
46	27	73.0	331	11	AAR05934	Bovine beta-1,4-ga
47	27	73.0	365	22	AAU01864	MAN_ATGT galactosy
48	27	73.0	372	21	AAG36697	Arabidopsis thalia
49	27	73.0	372	21	AAG37340	Arabidopsis thalia
50	27	73.0	372	21	AAG44321	Arabidopsis thalia

ALIGNMENTS

RESULT 1
AAY84686
ID AAY84686 standard; Protein; 2071 AA.
XX
AC AAY84686;
XX
DT 08-AUG-2000 (first entry)
XX
DE Amino acid sequence of AG003, a GTPase activating protein.
XX
KW GTP-binding protein; AG003; GTPase activating protein; fungal growth;
KW phosphatidylinositol-4 kinase; cytokenesis gene; fungal development;
KW fungicidal; filamentous fungi; plant pathogen; Septoria tritici;
KW Stagnospora nodorum; Magnaporthe grisea; human pathogen;
KW Candida albicans; Aspergillus fumigatus.
XX
OS Ashbya gossypii.
XX
PN WO200022133-A1.
XX
PD 20-APR-2000.
XX
PF 06-OCT-1999; 99WO-EP07501.
XX
PR 08-OCT-1998; 98US-0168804.
XX
PA (NOVS) NOVARTIS AG.
PA (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX

PI Gaffney TD, Wendland J, Dietrich F, Philippsen P, Goff SA;
XX
DR WPI; 2000-317988/27.
DR N-PSDB; AAA14503.
XX
PT Ashbya gossypii nucleotide and protein sequences useful for identifying
PT fungal growth inhibitors -
XX
PS Claim 2; Page 48-60; 93pp; English.
XX
CC The present sequence represents a putative GTPase activating protein,
CC designated AG003. The specification also describes putative GTP
CC binding protein genes, putative phosphatidylinositol-4 kinase
CC protein gene, and a putative cytokenesis gene. These genes are
CC essential for fungal growth and development. The proteins can be used
CC in methods to identify compounds that have fungicidal activity. Compounds
CC with fungicidal activity can be used for suppressing fungal growth,
CC especially of filamentous fungi. Fungi that can be suppressed include
CC plant pathogens (e.g. Septoria tritici, Stagnospora nodorum, and
CC Magnaporthe grisea) and human pathogens (e.g. Candida albicans, and
CC Aspergillus fumigatus).
XX
SQ Sequence 2071 AA;

Query Match 83.8%; Score 31; DB 21; Length 2071;
Best Local Similarity 85.7%; Pred. No. 3e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MNLVPMN 7
|||:|||
Db 718 mnlppmn 724

RESULT 2
AAB07839
ID AAB07839 standard; Protein; 237 AA.
XX
AC AAB07839;
XX
DT 14-NOV-2000 (first entry)
XX
DE Amino acid sequence of a soybean type IV glutathione-S-transferase.
XX
KW Soybean; glutathione-S-transferase; GST; detoxification;
KW xenobiotic compound; herbicide-tolerance; transgenic plant;
KW herbicide synergist.
XX
OS Glycine max.
XX
PN WO200047728-A2.
XX
PD 17-AUG-2000.
XX
PF 10-FEB-2000; 2000WO-US03347.
XX
PR 10-FEB-1999; 99US-0247373.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI McGonigle B, O'Keefe DP;
XX
DR WPI; 2000-549144/50.
DR N-PSDB; AAA59470.
XX
PT Soybean glutathione-S-transferase polypeptides and polynucleotides used
PT to produce herbicide tolerant transgenic plants and to screen for
PT inhibitors or substrates of the enzyme -
XX
PS Claim 4; Page 69-70; 84pp; English.
XX
CC The [illegible]ent sequence represents a soybean glutathione-S-transferase
CC [illegible]me. The enzyme is involved in the detoxification of
CC xenobiotic compounds in plants and seeds. The GST polynucleotides
CC and polypeptides are used for the production of herbicide-tolerant
CC transgenic plants, and for the development of screening assays to
CC identify GST inhibitors and substrates, which can be used as
CC herbicide synergists. GST Gene specific probes can be used in gene
CC identification methods. The recombinant GST enzymes can be used to
CC produce enzyme specific antibodies which are used to detect the
CC enzymes in situ in cells or in vitro in cell extracts.
XX
SQ Sequence 237 AA;

Query Match 78.4%; Score 29; DB 21; Length 237;
Best Local Similarity 71.4%; Pred. No. 79;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MNLVPMN 7
:||||:|
Db 55 inlvpin 61

RESULT 3
AAB03744
ID AAB03744 standard; Protein; 237 AA.
XX
AC AAB03744;
XX
DT 04-OCT-2000 (first entry)
XX
DE Clone se1.pk0017.f5 type IV GST protein sequence.
XX
KW Soybean; glutathione-S-transferase; GST; detoxify; herbicide; stress;
KW transgenic plant; tolerant; plant breeding.
XX
OS Glycine max.
XX
PN US6063570-A.
XX
PD 16-MAY-2000.
XX
PF 05-SEP-1997; 97US-0924747.
XX
PR 05-SEP-1997; 97US-0924747.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI McGonigle B, O'Keefe DP;
XX
DR WPI; 2000-375487/32.
DR N-PSDB; AAA53406.
XX
PT New Glutathione-S-Transferase enzymes and isolated nucleic acid
PT fragments encoding them, useful for detoxifying xenobiotic compounds in
PT plants and seeds, as well as in producing transgenic plants that are
PT herbicide-resistant -
XX
PS Claim 1; Column 65-66; 36pp; English.
XX
CC This sequence represents a Glutathione-S-Transferase (GST) protein
CC isolated from a soybean clone. The invention relates to isolated nucleic
CC acid fragments (see AAA53393-A53406) which encode soybean GST
CC polypeptides (AAB03731-B03744). GSTs are a family of enzymes which
CC catalyse the conjugation of glutathione, homoglutathione and other
CC glutathione-like analogues, to a large range of hydrophobic,
CC electrophilic compounds. GSTs have been implicated in the detoxification
CC of certain herbicides. The GST nucleotide sequences are useful in the
CC construction of herbicide-tolerant transgenic plants, plants that are
CC tolerant to a wide variety of stresses, or plants in which the GST
CC enzymes are present at higher or lower levels than they are normally. The
CC nucleic acid fragments are also useful as probes for genetically and
CC physically mapping the genes that they are part of, and as markers for
CC traits linked to expression of the enzymes. This will be useful in plant
CC breeding in order to develop lines with desired phenotypes or in the

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 6, 2002, 09:15:21 ; Search time 135.97 Seconds
(without alignments)
7.530 Million cell updates/sec

Title: US-09-897-042-14
Perfect score: 37
Sequence: 1 MNLVPMN 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : SPTREMBL_17:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	33	89.2	2451	12	Q9WJD4	Q9wjd4 venezuelan
2	32	86.5	336	2	Q9EVA1	Q9eva1 bacillus ce
3	32	86.5	346	2	Q9RVT6	Q9rvt6 deinococcus
4	32	86.5	492	5	Q27470	Q27470 caenorhabdi
5	32	86.5	561	2	Q9K6M4	Q9k6m4 bacillus ha
6	31	83.8	209	2	Q9L7X6	Q9l7x6 brucella ab
7	31	83.8	407	10	O82118	O82118 oryza sativ
8	31	83.8	1037	5	Q9NHF7	Q9nhf7 drosophila
9	31	83.8	2071	3	Q9HF75	Q9hf75 ashbya goss
10	30	81.1	346	2	Q9I2Q6	Q9i2q6 pseudomonas
11	30	81.1	369	2	O30580	O30580 bacillus su
12	30	81.1	619	10	Q9ZQ84	Q9zq84 arabidopsis
13	30	81.1	684	4	Q9UF31	Q9uf31 homo sapien
14	30	81.1	781	10	Q9ZUQ3	Q9zuq3 arabidopsis
15	30	81.1	827	2	O31506	O31506 bacillus su
16	30	81.1	1431	10	Q9C6V0	Q9c6v0 arabidopsis
17	30	81.1	2325	5	Q18104	Q18104 caenorhabdi
18	30	81.1	2475	12	Q08358	Q08358 african swi
19	29	78.4	113	12	Q66234	Q66234 cacao swoll
20	29	78.4	128	12	Q65288	Q65288 human adeno
21	29	78.4	237	10	Q9FQ95	Q9fq95 glycine max
22	29	78.4	280	5	Q9VI07	Q9vi07 drosophila
23	29	78.4	325	12	Q9QEE7	Q9qee7 indian citr
24	29	78.4	368	2	O86754	O86754 streptomyce
25	29	78.4	408	5	P91470	P91470 caenorhabdi
26	29	78.4	424	10	Q9ZUY3	Q9zuy3 arabidopsis
27	29	78.4	496	5	O16482	O16482 caenorhabdi
28	29	78.4	509	5	Q27482	Q27482 caenorhabdi
29	29	78.4	509	5	Q27499	Q27499 caenorhabdi
30	29	78.4	723	2	Q9AM59	Q9am59 acinetobact
31	29	78.4	732	5	O17619	O17619 caenorhabdi
32	29	78.4	933	10	Q9FPR3	Q9fpr3 arabidopsis
33	29	78.4	1879	12	Q9WIJ1	Q9wij1 venezuelan
34	29	78.4	2447	12	Q9WJD0	Q9wjd0 venezuelan
35	29	78.4	2455	12	Q9WJC8	Q9wjc8 venezuelan
36	29	78.4	2477	12	Q9WJD1	Q9wjd1 venezuelan
37	29	78.4	2478	12	Q9WJC6	Q9wjc6 venezuelan
38	29	78.4	2485	12	Q9WJD5	Q9wjd5 venezuelan
39	29	78.4	2485	12	Q9JFT5	Q9jft5 venezuelan
40	29	78.4	2485	12	Q9IR88	Q9ir88 venezuelan
41	29	78.4	2485	12	Q9ICG5	Q9icg5 venezuelan
42	29	78.4	2492	12	Q9IR87	Q9ir87 venezuelan
43	29	78.4	2493	12	Q66594	Q66594 venezuelan
44	29	78.4	2493	12	O90163	O90163 venezuelan
45	29	78.4	2493	12	Q66592	Q66592 venezuelan
46	29	78.4	2497	12	Q88978	Q88978 venezuelan
47	29	78.4	2498	12	Q86924	Q86924 aura virus.
48	29	78.4	2499	12	Q9WJC7	Q9wjc7 venezuelan
49	29	78.4	2501	12	Q9WJC9	Q9wjc9 venezuelan
50	29	78.4	2651	10	Q9FRR5	Q9frr5 arabidopsis

ALIGNMENTS

RESULT 1
Q9WJD4
ID Q9WJD4 PRELIMINARY; PRT; 2451 AA.
AC Q9WJD4;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE NONSTRUCTURAL POLYPROTEIN.
OS Venezuelan equine encephalitis virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
OC Alphavirus.
OX NCBI_TaxID=11036;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CABASSOU CAAR 508;
RX MEDLINE=99101297; PubMed=9886206;
RA Kinney R.M., Pfeffer M., Tsuchiya K.R., Chang G.J., Roehrig J.T.;
RT "Nucleotide sequences of the 26S mRNAs of the viruses defining the
RT Venezuelan equine encephalitis antigenic complex.";
RL Am. J. Trop. Med. Hyg. 59:952-964(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CABASSOU CAAR 508;
RA Kinney R.M., Pfeffer M., Meissner J.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF075259; AAD14566.1; -.
DR InterPro; IPR000606; Viral_helicase1.
DR InterPro; IPR002589; DUF27.
DR InterPro; IPR002620; Peptidase_C9.
DR Pfam; PF01443; Viral_helicase1; 1.
DR Pfam; PF01661; DUF27; 1.
DR Pfam; PF01707; Peptidase_C9; 1.
DR SMART; SM00506; A1pp; 1.
KW Polyprotein.
FT CHAIN 1 535 METHYLTRANSFERASE NSP1.
FT CHAIN 536 1329 NONSTRUCTURAL PROTEINASE NSP2.
FT CHAIN 1330 1844 NONSTRUCTURAL PROTEIN NSP3.

```
FT   CHAIN      1845   2451       RNA POLYMERASE NSP4.
SQ   SEQUENCE   2451 AA;  272975 MW;  2253DAEBF24F0383 CRC64;


  Query Match             89.2%;  Score 33;  DB 12;  Length 2451;
  Best Local Similarity   85.7%;  Pred. No. 1.5e+02;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 MNLVPMN 7
          :||||||
Db   1132 INLVPMN 1138


RESULT   2
Q9EVA1
ID   Q9EVA1      PRELIMINARY;      PRT;   336 AA.
AC   Q9EVA1;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CYTOTOXIN K.
GN   CYTK.
OS   Bacillus cereus.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
OX   NCBI_TaxID=1396;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   PubMed=11069652;
RA   Lund T., De Buyser M.L., Granum P.;
RT   "A new cytotoxin from Bacillus cereus that may cause necrotic
RT   enteritis.";
RL   Mol. Microbiol. 38:254-261(2000).
DR   EMBL; AJ277962; CAC08441.1; -.
DR   InterPro; IPR001340; Hemlysn_pore.
SQ   SEQUENCE   336 AA;  36953 MW;  54C1148EAE1B709E CRC64;


  Query Match             86.5%;  Score 32;  DB 2;  Length 336;
  Best Local Similarity   100.0%;  Pred. No. 32;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 NLVPMN 7
          ||||||
Db    248 NLVPMN 253


RESULT   3
Q9RVT6
ID   Q9RVT6      PRELIMINARY;      PRT;   346 AA.
AC   Q9RVT6;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   CONSERVED HYPOTHETICAL PROTEIN.
GN   DR0936.
OS   Deinococcus radiodurans.
OC   Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX   NCBI_TaxID=1299;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=R1;
RX   MEDLINE=20036896; PubMed=10567266;
RA   White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA   Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA   Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA   Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA   Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA   Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA   Fraser C.M.;
RT   "Genome sequence of the radioresistant bacterium Deinococcus
RT   radiodurans R1.";
RL   Science 286:1571-1577(1999).
DR   EMBL; AE001946; AAF10513.1; -.
DR   TIGR; DR0936; -.
KW   Complete proteome.
SQ   SEQUENCE   346 AA;  37979 MW;  64506136A45F9C43 CRC64;


  Query Match             86.5%;  Score 32;  DB 2;  Length 346;
  Best Local Similarity   71.4%;  Pred. No. 33;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 MNLVPMN 7
          :||:|||
Db    279 VNLIPMN 285


RESULT   4
Q27470
ID   Q27470      PRELIMINARY;      PRT;   492 AA.
AC   Q27470;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   SIMILAR TO CYTOCHROME P450 (EC 1.14.14.1).
GN   C12D5.7.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., Mcmurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA   Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Miller N., Stellyes L., Bradshaw H.;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RL   Submitted (APR-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR   EMBL; U55365; AAA98571.1; -.
DR   InterPro; IPR001128; Cyt_P450.
DR   Pfam; PF00067; p450; 1.
DR   PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW   Electron transport; Heme; Membrane; Monooxygenase; Oxidoreductase.
FT   BINDING     436    436       HEME (BY SIMILARITY).
SQ   SEQUENCE   492 AA;  56770 MW;  52B71FE98F0B06C4 CRC64;


  Query Match             86.5%;  Score 32;  DB 5;  Length 492;
  Best Local Similarity   100.0%;  Pred. No. 47;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 NLVPMN 7
          ||||||
Db    361 NLVPMN 366
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 6, 2002, 09:16:56 ; Search time 68.88 Seconds
(without alignments)
3.726 Million cell updates/sec

Title: US-09-897-042-14
Perfect score: 37
Sequence: 1 MNLVPMN 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID		Description
1	31	83.8	210	1	CLPP_CAUCR	O87706	caulobacter
2	29	78.4	142	1	SECB_BUCAI	P57161	buchnera ap
3	29	78.4	364	1	YS80_MYCTU	Q10806	mycobacteri
4	29	78.4	575	1	PUT2_YEAST	P07275	saccharomyc
5	29	78.4	679	1	PAN3_YEAST	P36102	saccharomyc
6	29	78.4	2485	1	POLN_EEVV3	P36327	venezuelan
7	29	78.4	2492	1	POLN_EEVVP	P36328	venezuelan
8	29	78.4	2492	1	POLN_EEVVT	P27282	venezuelan
9	28	75.7	211	1	LIPB_BUCAI	P57356	buchnera ap
10	28	75.7	215	1	FIMI_ECOLI	P39264	escherichia
11	28	75.7	317	1	Y402_RICPR	Q9zdc9	rickettsia
12	28	75.7	363	1	YLON_BACSU	O34617	bacillus su
13	28	75.7	389	1	NCAP_CVH22	P15130	human coron
14	28	75.7	535	1	VG50_HSVSA	Q01012	herpesvirus
15	28	75.7	536	1	CCA4_DROME	Q9ve00	drosophila
16	28	75.7	596	1	MPIP_SCHPO	P06652	schizosacch
17	28	75.7	617	1	ASMA_ECOLI	P28249	escherichia
18	28	75.7	905	1	GYRA_RICPR	P41080	rickettsia
19	28	75.7	2512	1	POLN_SINDV	P03317	sindbis vir
20	28	75.7	2514	1	POLN_SINDO	P27283	sindbis vir
21	27	73.0	112	1	OLF8_MOUSE	Q60892	mus musculu
22	27	73.0	187	1	KIP2_HUMAN	O75838	homo sapien
23	27	73.0	187	1	KIP2_MOUSE	Q9z309	mus musculu
24	27	73.0	298	1	Y680_HAEIN	Q57389	haemophilus
25	27	73.0	317	1	OLF2_RAT	P23268	rattus norv
26	27	73.0	345	1	TOB2_MOUSE	Q9jm55	mus musculu
27	27	73.0	379	1	Y2A6_PSEAE	Q51385	pseudomonas
28	27	73.0	395	1	SYW_AQUAE	O67115	aquifex aeo
29	27	73.0	400	1	B4G1_HUMAN	P15291	h beta-1,4-
30	27	73.0	402	1	B4G1_BOVIN	P08037	b beta-1,4-
31	27	73.0	406	1	YMX2_CAEEL	P34510	caenorhabdi
32	27	73.0	472	1	YWIE_CAEEL	Q23525	caenorhabdi
33	27	73.0	493	1	CPE1_MOUSE	Q05421	mus musculu
34	27	73.0	493	1	CPE1_RAT	P05182	rattus norv
35	27	73.0	501	1	CPJ5_MOUSE	O54749	mus musculu
36	27	73.0	505	1	CATA_METBA	O93662	methanosarc
37	27	73.0	533	1	YADC_SCHPO	Q09837	schizosacch
38	27	73.0	553	1	5NTD_VIBCH	Q9kq30	vibrio chol
39	27	73.0	569	1	YE59_MYCPN	P75324	mycoplasma
40	27	73.0	671	1	KPC1_HUMAN	P05771	homo sapien
41	27	73.0	671	1	KPC1_RABIT	P05772	oryctolagus
42	27	73.0	671	1	KPC1_RAT	P04410	rattus norv
43	27	73.0	673	1	KPC2_BOVIN	P05126	bos taurus
44	27	73.0	673	1	KPC2_HUMAN	P05127	homo sapien
45	27	73.0	673	1	KPC2_MOUSE	P04411	mus musculu
46	27	73.0	673	1	KPC2_RABIT	P05773	oryctolagus
47	27	73.0	688	1	ENV_MMTVB	P10259	mouse mamma
48	27	73.0	1012	1	POLG_IBDVO	P27276	avian infec
49	26	70.3	91	1	YURS_BACSU	O32160	bacillus su
50	26	70.3	119	1	Y167_UREPA	Q9pqx6	ureaplasma

ALIGNMENTS

RESULT 1
CLPP_CAUCR
ID CLPP_CAUCR STANDARD; PRT; 210 AA.
AC O87706;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ATP-DEPENDENT CLP PROTEASE PROTEOLYTIC SUBUNIT (EC 3.4.21.92)
DE (ENDOPEPTIDASE CLP).
GN CLPP OR CC1963.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=69394;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CB15N / NA1000;
RX MEDLINE=98429489; PubMed=9755166;
RA Jenal U., Fuchs T.;
RT "An essential protease involved in bacterial cell-cycle control.";
RL EMBO J. 17:5658-5669(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.K.R., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Shapiro L., Venter J.C., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
CC -!- FUNCTION: CLP CLEAVES PEPTIDES IN VARIOUS PROTEINS IN A PROCESS
CC THAT REQUIRES ATP HYDROLYSIS. CLP MAY BE RESPONSIBLE FOR A FAIRLY
CC GENERAL AND CENTRAL HOUSEKEEPING FUNCTION RATHER THAN FOR THE
CC DEGRADATION OF SPECIFIC SUBSTRATES (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: HYDROLYSIS OF PROTEINS TO SMALL PEPTIDES IN
CC THE PRESENCE OF ATP AND MAGNESIUM. ALPHA-CASEIN IS THE USUAL TEST
CC SUBSTRATE. IN THE ABSENCE OF ATP, ONLY OLIGOPEPTIDES SHORTER THAN
CC FIVE RESIDUES ARE CLEAVED (SUCH AS SUCCINYL-LEU-TYR-|-NHMEC; AND
CC LEU-TYR-LEU-|-TYR-TRP, IN WHICH THE CLEAVAGE OF THE -TYR-|-LEU-
CC AND -TYR-|-TRP-BOND ALSO OCCURS).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S14; ALSO KNOWN AS CLPP
CC FAMILY.
CC ---

```
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AJ010321; CAA09090.1; -.
DR   EMBL; AE005869; AAK23938.1; -.
DR   MEROPS; S14.003; -.
DR   TIGR; CC1963; -.
DR   InterPro; IPR001907; CLP_protease.
DR   Pfam; PF00574; CLP_protease; 1.
DR   PRINTS; PR00127; CLPPROTEASEP.
DR   PROSITE; PS00381; CLP_PROTEASE_SER; 1.
DR   PROSITE; PS00382; CLP_PROTEASE_HIS; 1.
KW   Hydrolase; Serine protease; Complete proteome.
FT   ACT_SITE    107    107       PROBABLE.
FT   ACT_SITE    132    132       PROBABLE.
SQ   SEQUENCE   210 AA;  23120 MW;  CC2421D44C1E7E59 CRC64;


  Query Match             83.8%;  Score 31;  DB 1;  Length 210;
  Best Local Similarity   100.0%;  Pred. No. 8.5;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MNLVPM 6
           ||||||
Db      10 MNLVPM 15


RESULT   2
SECB_BUCAI
ID   SECB_BUCAI     STANDARD;      PRT;   142 AA.
AC   P57161;
DT   20-AUG-2001 (Rel. 40, Created)
DT   20-AUG-2001 (Rel. 40, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   PROTEIN-EXPORT PROTEIN SECB.
GN   SECB OR BU053.
OS   Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OS   symbiotic bacterium).
OC   Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX   NCBI_TaxID=118099;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TOKYO 1998;
RX   MEDLINE=20445173; PubMed=10993077;
RA   Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT   "Genome sequence of the endocellular bacterial symbiont of aphids
RT   Buchnera sp. APS.";
RL   Nature 407:81-86(2000).
CC   -!- FUNCTION: THIS IS ONE OF THE PROTEINS REQUIRED FOR THE NORMAL
CC       EXPORT OF ENVELOPE PROTEINS OUT OF THE CELL CYTOPLASM; IT MAY
CC       BE INVOLVED IN THE INITIATION OF THE EXPORTING PROCESS, BY BINDING
CC       TO THE NASCENT POLYPEPTIDE VIA A SIGNAL SEQUENCE, MAINTAINING
CC       A STABLE AND PRE-TRANSLOCATION CONFORMATION (BY SIMILARITY).
CC   -!- SUBUNIT: PART OF THE PROKARYOTIC PROTEIN TRANSLOCATION APPARATUS
CC       WHICH COMPRISE SECA, SECB, SECE, SECF, SECG AND SECY
CC       (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE SECB FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AP001118; BAB12776.1; -.
DR   InterPro; IPR003708; SecB.
DR   Pfam; PF02556; SecB; 1.
KW   Protein transport; Translocation; Complete proteome.
SQ   SEQUENCE   142 AA;  16571 MW;  C81D1176607D91D9 CRC64;


  Query Match             78.4%;  Score 29;  DB 1;  Length 142;
  Best Local Similarity   71.4%;  Pred. No. 16;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 MNLVPMN 7
           ||| |:|
Db     126 MNLAPIN 132


RESULT   3
YS80_MYCTU
ID   YS80_MYCTU     STANDARD;      PRT;   364 AA.
AC   Q10806; Q10805;
DT   01-OCT-1996 (Rel. 34, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   HYPOTHETICAL 29.6 KDA PROTEIN RV2880C/RV2879C.
GN   RV2880C/RV2879C OR MT2947 OR MTCY274.11C/MTCY274.10C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE UPF0063 FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; Z74024; CAA98356.1; ALT_FRAME.
DR   EMBL; Z74024; CAA98355.1; ALT_FRAME.
DR   EMBL; AE007118; AAK47272.1; -.
DR   TIGR; MT2947; -.
DR   TubercuList; Rv2879c; -.
DR   TubercuList; Rv2880c; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   364 AA;  39255 MW;  74351614D3484A32 CRC64;


  Query Match             78.4%;  Score 29;  DB 1;  Length 364;
  Best Local Similarity   57.1%;  Pred. No. 40;
  Matches    4;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: January 6, 2002, 09:13:46 ; Search time 77.97 Seconds
(without alignments)
6.839 Million cell updates/sec

Title: US-09-897-042-14
Perfect score: 37
Sequence: 1 MNLVPMN 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	32	86.5	346	2	F75457	conserved hypothet
2	32	86.5	492	2	T30066	hypothetical prote
3	32	86.5	561	2	A84113	transposase (12) B
4	31	83.8	407	2	JE0113	zinc-finger protei
5	30	81.1	346	2	A83415	hypothetical prote
6	30	81.1	619	2	A84449	hypothetical prote
7	30	81.1	684	2	T43452	hypothetical prote
8	30	81.1	781	2	H84501	En/Spm-like transp
9	30	81.1	827	2	A69793	conserved hypothet
10	30	81.1	2325	2	T15566	hypothetical prote
11	30	81.1	2475	2	S35307	polyprotein pp220
12	29	78.4	128	2	S33949	early E3B 14.7K pr
13	29	78.4	142	2	H84935	protein-export pro
14	29	78.4	189	2	B70924	conserved hypothet
15	29	78.4	368	2	T35453	hypothetical prote
16	29	78.4	408	2	T29949	hypothetical prote
17	29	78.4	424	2	D84677	hypothetical prote
18	29	78.4	509	2	T29849	hypothetical prote
19	29	78.4	509	2	T29303	hypothetical prote
20	29	78.4	575	1	RDBYC	1-pyrroline-5-carb
21	29	78.4	679	2	S37842	hypothetical prote
22	29	78.4	732	2	T19570	hypothetical prote
23	29	78.4	1015	2	T00726	probable serine/th
24	29	78.4	2492	1	MNWVTD	nonstructural poly
25	29	78.4	2492	1	A44213	nonstructural poly
26	29	78.4	2492	1	C44213	nonstructural poly
27	28	75.7	208	2	A82712	endopeptidase Clp
28	28	75.7	211	2	B84961	lipoate-protein li
29	28	75.7	215	2	S56540	fimI protein - Esc
30	28	75.7	215	2	E86129	fimbrial protein [
31	28	75.7	220	2	T43857	hypothetical prote
32	28	75.7	317	2	A71698	hypothetical prote
33	28	75.7	330	2	A86355	hypothetical prote
34	28	75.7	343	2	G72218	conserved hypothet
35	28	75.7	362	2	B83963	hypothetical prote
36	28	75.7	363	2	F69878	conserved hypothet
37	28	75.7	389	1	VHIH2E	nucleocapsid prote
38	28	75.7	389	2	S08031	nucleocapsid prote
39	28	75.7	514	2	E86738	multidrug resistan
40	28	75.7	533	2	F82237	conserved hypothet
41	28	75.7	535	2	C36811	hypothetical prote
42	28	75.7	535	2	A37994	RF1 protein - saim
43	28	75.7	596	1	S62407	protein-tyrosine-p
44	28	75.7	617	2	G64972	yegA protein precu
45	28	75.7	905	2	H71731	DNA gyrase chain A
46	28	75.7	2512	1	MNWVS	nonstructural poly
47	28	75.7	2514	1	MNWV82	nonstructural poly
48	27	73.0	44	2	I37286	olfactory receptor
49	27	73.0	44	2	E37286	olfactory receptor
50	27	73.0	96	2	T18043	hypothetical prote

ALIGNMENTS

RESULT 1
F75457
conserved hypothetical protein - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C;Accession: F75457
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J , M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896
A;Accession: F75457
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-346 <WHI>
A;Cross-references: GB:AE001946; GB:AE000513; NID:g6458655; PIDN:AAF10513.1; PID:g645
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR0936
A;Map position: 1
C;Superfamily: conserved hypothetical protein HI0365

Query Match 86.5%; Score 32; DB 2; Length 346;
Best Local Similarity 71.4%; Pred. No. 13;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MNLVPMN 7
:||:|||
Db 279 VNLIPMN 285

RESULT 2
T30066
hypothetical protein C12D5.7 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 16-Feb-2001
C;Accession: T30066
R;Miller, N.; Stellyes, L.; Bradshaw, H.
submitted to the EMBL Data Library, April 1996
A;Description: The sequence of C. elegans cosmid C12D5.
A;Reference number: Z20729
A;Accession: T30066
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-492 <MIL>

```
A;Cross-references: EMBL:U55365; PIDN:AAA98571.1; GSPDB:GN00023; CESP:C12D5.7
A;Experimental source: strain Bristol N2; clone C12D5
C;Genetics:
A;Gene: CESP:C12D5.7
A;Map position: 5
A;Introns: 28/1; 166/3; 231/2; 354/3
C;Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology
C;Keywords: heme; iron; metalloprotein
F;296-458/Domain: cytochrome P450 homology <P45>
F;436/Binding site: heme iron (Cys) (axial ligand) #status predicted


  Query Match             86.5%;  Score 32;  DB 2;  Length 492;
  Best Local Similarity   100.0%;  Pred. No. 19;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 NLVPMN 7
           ||||||
 b     361 NLVPMN 366


RESULT   3
A84113
transposase (12) BH3705 [imported] - Bacillus halodurans (strain C-125)
C;Species: Bacillus halodurans
C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 31-Dec-2000
C;Accession: A84113
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A;Reference number: A83650; MUID:20263314
A;Accession: A84113
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-561 <STO>
A;Cross-references: GB:AP001519; GB:BA000004; NID:g10176109; PIDN:BAB07424.1; GSPDB:GN00
A;Experimental source: strain C-125
C;Genetics:
A;Gene: BH3705


  Query Match             86.5%;  Score 32;  DB 2;  Length 561;
  Best Local Similarity   100.0%;  Pred. No. 22;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 NLVPMN 7
           ||||||
Db     491 NLVPMN 496


RESULT   4
JE0113
zinc-finger protein S3574 [imported] - rice
C;Species: Oryza sativa (rice)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 20-Apr-2000
C;Accession: JE0113
R;Song, J.; Yamamoto, K.; Shomura, A.; Itadani, H.; Zhong, H.S.; Yano, M.; Sasaki, T.
DNA Res. 5, 95-101, 1998
A;Title: Isolation and mapping of a family of putative zinc-finger protein cDNAs from ri
A;Reference number: JE0113; MUID:98344140
A;Accession: JE0113
A;Molecule type: mRNA
A;Residues: 1-407 <SON>
A;Cross-references: DDBJ:AB001888; NID:g3618319; PIDN:BAA33206.1; PID:d1034172; PID:g361


  Query Match             83.8%;  Score 31;  DB 2;  Length 407;
  Best Local Similarity   71.4%;  Pred. No. 26;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MNLVPMN 7
           |||:|:|
Db     177 MNLLPLN 183


RESULT   5
A83415
hypothetical protein PA1839 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: A83415
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A;Reference number: A82950; MUID:20437337
A;Accession: A83415
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-346 <STO>
A;Cross-references: GB:AE004610; GB:AE004091; NID:g9947825; PIDN:AAG05228.1; GSPDB:GN
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA1839
C;Superfamily: conserved hypothetical protein HI0365


  Query Match             81.1%;  Score 30;  DB 2;  Length 346;
  Best Local Similarity   71.4%;  Pred. No. 37;
  Matches    5;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 MNLVPMN 7
           |||:| |
Db     276 MNLIPYN 282


RESULT   6
A84449
hypothetical protein At2g03480 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: A84449
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487
A;Accession: A84449
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-619 <STO>
A;Cross-references: GB:AE002093; NID:g4335751; PIDN:AAD17428.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g03480
A;Map position: 2


  Query Match             81.1%;  Score 30;  DB 2;  Length 619;
  Best Local Similarity   71.4%;  Pred. No. 71;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MNLVPMN 7
           ||:||:|
Db     500 MNVVPVN 506


RESULT   7
T43452
hypothetical protein DKFZp434E2118.1 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 03-Nov-2000
C;Accession: T43452
R;Ottenwaelder, B.; Obermaier, B.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, December 1999
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: January 6, 2002, 09:48:12 ; Search time 121.65 Seconds
(without alignments)
44.681 Million cell updates/sec

Title: US-09-897-042-13
Perfect score: 24
Sequence: 1 atgaacttggtaccgatgaactaa 24

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	16.6	69.2	1452	2	US-08-966-316-12	Sequence 12, Appl
c	2	16.6	69.2	1452	2	US-08-966-316-12	Sequence 12, Appl
c	3	16.4	68.3	2097	6	5460961-4	Patent No. 5460961
	4	16.4	68.3	2261	6	5460961-1	Patent No. 5460961
	5	16.2	67.5	658	4	US-09-328-111-816	Sequence 816, App
	6	16.2	67.5	1269	1	US-08-265-429A-4	Sequence 4, Appli
	7	16.2	67.5	1269	5	PCT-US95-09069-4	Sequence 4, Appli
	8	16.2	67.5	1401	4	US-08-952-457-1	Sequence 1, Appli
	9	16.2	67.5	3404	1	US-08-265-429A-1	Sequence 1, Appli
	10	16.2	67.5	3404	5	PCT-US95-09069-1	Sequence 1, Appli
c	11	16	66.7	752	4	US-08-998-416-568	Sequence 568, App
	12	16	66.7	784	3	US-08-961-083-163	Sequence 163, App
	13	16	66.7	3135	4	US-09-321-276-1	Sequence 1, Appli
	14	16	66.7	3135	4	US-09-321-276-3	Sequence 3, Appli
	15	16	66.7	3135	4	US-08-916-481-1	Sequence 1, Appli
	16	16	66.7	3135	4	US-08-916-481-4	Sequence 4, Appli
	17	15.6	65.0	1608	1	US-07-621-670-2	Sequence 2, Appli
c	18	15.6	65.0	1800	3	US-09-039-773A-1	Sequence 1, Appli
c	19	15.6	65.0	11970	4	US-09-345-217-1	Sequence 1, Appli
c	20	15.6	65.0	152331	3	US-09-128-155-16	Sequence 16, Appl
c	21	15.2	63.3	2574	4	US-09-142-529-2	Sequence 2, Appli
c	22	15.2	63.3	3828	5	PCT-US93-10500-1	Sequence 1, Appli
c	23	15.2	63.3	4190	2	US-08-488-706-3	Sequence 3, Appli
c	24	15.2	63.3	4517	5	PCT-US93-06251-83	Sequence 83, Appl
c	25	15.2	63.3	7721	3	US-08-772-270A-14	Sequence 14, Appl
	26	15	62.5	208	1	US-08-480-552-17	Sequence 17, Appl
	27	15	62.5	208	3	US-08-929-208-17	Sequence 17, Appl
	28	15	62.5	1760	3	US-09-082-310-4	Sequence 4, Appli
	29	15	62.5	2400	6	5256558-6	Patent No. 5256558
	30	15	62.5	3254	1	US-08-162-809-15	Sequence 15, Appl
	31	15	62.5	10660	2	US-08-267-803B-8	Sequence 8, Appli
	32	15	62.5	10660	4	US-09-041-886-16	Sequence 16, Appl
c	33	14.8	61.7	1577	2	US-08-717-587A-1	Sequence 1, Appli
c	34	14.8	61.7	1577	3	US-08-883-610A-1	Sequence 1, Appli
c	35	14.8	61.7	1577	4	US-08-936-094A-1	Sequence 1, Appli
c	36	14.8	61.7	4403765	4	US-09-103-840A-2	Sequence 2, Appli
c	37	14.8	61.7	4411529	4	US-09-103-840A-1	Sequence 1, Appli
c	38	14.6	60.8	445	1	US-08-592-126-67	Sequence 67, Appl
c	39	14.6	60.8	467	4	US-08-927-219-133	Sequence 133, App
	40	14.6	60.8	602	4	US-08-998-416-52	Sequence 52, Appl
c	41	14.6	60.8	748	4	US-08-998-416-263	Sequence 263, App
c	42	14.6	60.8	751	4	US-08-998-416-399	Sequence 399, App
c	43	14.6	60.8	881	1	US-08-631-200-12	Sequence 12, Appl
c	44	14.6	60.8	881	1	US-08-829-553-12	Sequence 12, Appl
c	45	14.6	60.8	881	2	US-08-922-267A-12	Sequence 12, Appl

ALIGNMENTS

```
RESULT   1
US-08-966-316-12
; Sequence 12, Application US/08966316
; Patent No. 5932445
;  GENERAL INFORMATION:
;    APPLICANT:  Lal, Preeti
;    APPLICANT:  Au-Young, Janice
;    APPLICANT:  Reddy, Roopa
;    APPLICANT:  Murry, Lynn E.
;    APPLICANT:  Mathur, Preete
;    TITLE OF INVENTION:  SIGNAL PEPTIDE - CONTAINING PROTEINS
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/966,316
;      FILING DATE:  Herewith
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0424 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-855-0555
;      TELEFAX:  650-845-4166
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1452 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  EOSIHET02
;      CLONE:  318060
US-08-966-316-12
```

```
  Query Match             69.2%;  Score 16.6;  DB 2;  Length 1452;
  Best Local Similarity   82.6%;  Pred. No. 32;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 atgaacttggtaccgatgaacta 23
          ||| ||||  ||| |||||||||
Db   1027 ATGCACTTCCTACAGATGAACTA 1049


RESULT   2
US-08-966-316-12/c
; Sequence 12, Application US/08966316
; Patent No. 5932445
;  GENERAL INFORMATION:
;    APPLICANT:  Lal, Preeti
;    APPLICANT:  Au-Young, Janice
;    APPLICANT:  Reddy, Roopa
;    APPLICANT:  Murry, Lynn E.
;    APPLICANT:  Mathur, Preete
;    TITLE OF INVENTION:  SIGNAL PEPTIDE - CONTAINING PROTEINS
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/966,316
;      FILING DATE:  Herewith
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0424 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-855-0555
;      TELEFAX:  650-845-4166
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1452 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  EOSIHET02
;      CLONE:  318060
US-08-966-316-12


  Query Match             69.2%;  Score 16.6;  DB 2;  Length 1452;
  Best Local Similarity   82.6%;  Pred. No. 32;
  Matches   19;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 atgaacttggtaccgatgaacta 23
          ||| ||||  ||| |||||||||
Db    410 ATGCACTTCCTACAGATGAACTA 388


RESULT   3
5460961-4/c
;Patent No. 5460961
;    APPLICANT: DEBY, CAROL;PINCEMAIL, JOEL;BOLLEN, ALEX
;    TITLE OF INVENTION: HUMAN MYELOPEROXIDASE AND ITS
;THERAPEUTIC APPLICATION
;    NUMBER OF SEQUENCES: 13
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/641,678
;      FILING DATE: 16-JAN-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: 460,931
;      FILING DATE: 14-FEB-1990
;SEQ ID NO:4:
;      LENGTH: 2097
5460961-4


  Query Match             68.3%;  Score 16.4;  DB 6;  Length 2097;
  Best Local Similarity   94.4%;  Pred. No. 41;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 tgaacttggtaccgatga 19
          ||||||| ||||||||||
Db   1864 TGAACTGGGTACCGATGA 1847


RESULT   4
5460961-1/c
;Patent No. 5460961
;    APPLICANT: DEBY, CAROL;PINCEMAIL, JOEL;BOLLEN, ALEX
;    TITLE OF INVENTION: HUMAN MYELOPEROXIDASE AND ITS
;THERAPEUTIC APPLICATION
;    NUMBER OF SEQUENCES: 13
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/641,678
;      FILING DATE: 16-JAN-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: 460,931
;      FILING DATE: 14-FEB-1990
;SEQ ID NO:1:
;      LENGTH: 2261
5460961-1


  Query Match             68.3%;  Score 16.4;  DB 6;  Length 2261;
  Best Local Similarity   94.4%;  Pred. No. 41;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 tgaacttggtaccgatga 19
          ||||||| ||||||||||
Db   2016 TGAACTGGGTACCGATGA 1999


RESULT   5
US-09-328-111-816
; Sequence 816, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
;  APPLICANT: Endege, Wilson O.
;  APPLICANT: Steinmann, Kathleen E.
;  APPLICANT: Astle, Jon H.
;  APPLICANT: Burgess, Christopher C.
;  APPLICANT: Bushnell, Steven E.
;  APPLICANT: Carroll III, Eddie
;  APPLICANT: Catino, Theodore J.
;  APPLICANT: Derti, Adnan
;  APPLICANT: Ford, Donna M.
;  APPLICANT: Lewis, Marcia E.
;  APPLICANT: Monahan, John E.
;  APPLICANT: Schlegel, Robert
;  TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
;  TITLE OF INVENTION: PRODUCTS
;  FILE REFERENCE: CCD-257 (US)
;  CURRENT APPLICATION NUMBER: US/09/328,111
;  CURRENT FILING DATE: 1999-06-08
```

```
CC   This invention relates to nucleotide sequences AAF29337 - AAF29339 which
CC   encode atopy related proteins. The invention includes vectors containing
CC   any of the DNA sequences, a drug containing any of the DNA sequences, and
CC   a cell transformed by the vector. The atopy genes can be used in the
CC   diagnosis and treatment of allergic diseases, atopic dermatitis and
CC   asthma.
XX
SQ   Sequence 33030 BP; 9416 A; 6081 C; 6495 G; 10501 T; 537 other;


  Query Match             73.3%;  Score 17.6;  DB 22;  Length 33030;
  Best Local Similarity   58.3%;  Pred. No. 45;
  Matches   14;  Conservative    8;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 atgaacttggtaccgatgaactaa 24
          ||||||||||:::|::: |: :|
Db    737 atgaacttggtwmmgrwkkamama 760


RESULT   2
AAC49803
ID   AAC49803 standard; DNA; 573 BP.
XX
AC   AAC49803;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 62494.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
PR   01-JUL-1999;   99US-0142154.
PR   02-JUL-1999;   99US-0142055.
PR   06-JUL-1999;   99US-0142390.
PR   08-JUL-1999;   99US-0142803.
PR   09-JUL-1999;   99US-0142920.
PR   12-JUL-1999;   99US-0142977.
PR   13-JUL-1999;   99US-0143542.
PR   14-JUL-1999;   99US-0143624.
PR   15-JUL-1999;   99US-0144005.
PR   16-JUL-1999;   99US-0144085.
PR   16-JUL-1999;   99US-0144086.
PR   19-JUL-1999;   99US-0144325.
PR   19-JUL-1999;   99US-0144331.
PR   19-JUL-1999;   99US-0144332.
PR   19-JUL-1999;   99US-0144333.
PR   19-JUL-1999;   99US-0144334.
PR   19-JUL-1999;   99US-0144335.
PR   20-JUL-1999;   99US-0144352.
PR   20-JUL-1999;   99US-0144632.
PR   20-JUL-1999;   99US-0144884.
PR   21-JUL-1999;   99US-0144814.
PR   21-JUL-1999;   99US-0145086.
PR   21-JUL-1999;   99US-0145088.
PR   22-JUL-1999;   99US-0145085.
PR   22-JUL-1999;   99US-0145087.
PR   22-JUL-1999;   99US-0145089.
PR   22-JUL-1999;   99US-0145192.
PR   23-JUL-1999;   99US-0145145.
PR   23-JUL-1999;   99US-0145218.
PR   23-JUL-1999;   99US-0145224.
PR   26-JUL-1999;   99US-0145276.
PR   27-JUL-1999;   99US-0145913.
PR   27-JUL-1999;   99US-0145918.
PR   27-JUL-1999;   99US-0145919.
PR   28-JUL-1999;   99US-0145951.
PR   02-AUG-1999;   99US-0146386.
PR   02-AUG-1999;   99US-0146388.
PR   02-AUG-1999;   99US-0146389.
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: January 6, 2002, 09:48:47 ; Search time 256.95 Seconds
(without alignments)
80.077 Million cell updates/sec

Title: US-09-897-042-13
Perfect score: 24
Sequence: 1 atgaacttggtaccgatgaactaa 24

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_1101:*
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	17.6	73.3	33030	22	AAF29337	Atopy related gene
	2	16.6	69.2	573	21	AAC49803	Arabidopsis thalia
	3	16.6	69.2	823	21	AAC35318	Arabidopsis thalia
	4	16.6	69.2	1452	20	AAX61284	Human signal pepti
c	5	16.6	69.2	1452	20	AAX61284	Human signal pepti
c	6	16.4	68.3	237	19	AAV11521	Recombinant MPO DN
c	7	16.4	68.3	257	22	AAI25101	Probe #15034 for g
c	8	16.4	68.3	257	22	AAI50972	Probe #19658 used
c	9	16.4	68.3	444	19	AAV11515	Recombinant MPO DN
c	10	16.4	68.3	615	19	AAV11514	Recombinant MPO DN
c	11	16.4	68.3	702	19	AAV11509	Recombinant MPO DN
c	12	16.4	68.3	1386	19	AAV11507	Recombinant MPO DN
c	13	16.4	68.3	2260	11	AAQ03118	HindIII/HpaII frag
c	14	16.4	68.3	3215	18	AAT66437	Myeloperoxidase co
	15	16.2	67.5	424	21	AAC31939	Human secreted pro
c	16	16.2	67.5	516	21	AAC54535	Arabidopsis thalia
c	17	16.2	67.5	586	21	AAC34988	Arabidopsis thalia
	18	16.2	67.5	658	21	AAZ80732	Human colon cancer
	19	16.2	67.5	1269	17	AAT15596	Thermostable cellu
	20	16.2	67.5	1344	18	AAT73169	Pseudomonas putida
	21	16.2	67.5	1349	18	AAT73177	Pseudomonas putida
	22	16.2	67.5	1350	18	AAT73176	Pseudomonas putida
	23	16.2	67.5	1366	18	AAT73174	Pseudomonas putida
	24	16.2	67.5	1374	18	AAT73173	Pseudomonas putida
	25	16.2	67.5	1395	18	AAT73172	Pseudomonas putida
	26	16.2	67.5	1401	18	AAT73167	Pseudomonas putida
	27	16.2	67.5	1412	18	AAT73175	Pseudomonas putida
	28	16.2	67.5	1417	18	AAT73171	Pseudomonas putida
	29	16.2	67.5	1441	18	AAT73170	Pseudomonas putida
c	30	16.2	67.5	1967	21	AAC78171	Human cancer assoc
c	31	16.2	67.5	3239	22	AAD05543	Human secreted pro
c	32	16.2	67.5	3265	22	AAD08193	Human secreted pro
c	33	16.2	67.5	3265	22	AAD05545	Human secreted pro
c	34	16.2	67.5	3296	22	AAD12580	Human protein havi
c	35	16.2	67.5	3320	22	AAS23196	DNA encoding novel
c	36	16.2	67.5	3339	22	AAD05503	Human secreted pro
c	37	16.2	67.5	3339	22	AAD05544	Human secreted pro
c	38	16.2	67.5	3349	21	AAA16628	Human secreted pro
c	39	16.2	67.5	3362	22	AAF93802	Human cDNA encodin
	40	16.2	67.5	3404	17	AAT15595	Thermostable cellu
	41	16.2	67.5	3503	19	AAV07164	Thermomonospora fu
	42	16	66.7	146	22	AAI51261	Probe #19947 used
c	43	16	66.7	285	21	AAC94537	Cat flea hindgut a
	44	16	66.7	317	20	AAV88731	EST clone HJ253.
c	45	16	66.7	360	14	AAQ61040	Human brain Expres

ALIGNMENTS

RESULT 1
AAF29337
ID AAF29337 standard; DNA; 33030 BP.
XX
AC AAF29337;
XX
DT 20-APR-2001 (first entry)
XX
DE Atopy related gene sequence SEQ ID 1.
XX
KW Atopy; human; allergic disease; atopic dermatitis; asthma; ds.
XX
OS Homo sapiens.
XX
PN WO200104302-A1.
XX
PD 18-JAN-2001.
XX
PF 10-JUL-2000; 2000WO-JP04599.
XX
PR 09-JUL-1999; 99JP-0196517.
XX
PA (MITU) MITSUBISHI CHEM CORP.
XX
PI Adra CN, Shirakawa T, Hopkin JM;
XX
DR WPI; 2001-138350/14.
XX
PT New atopy gene with domain containing HTm4 gene in eleventh chromosome
PT q13.1 domain identified and isolated by PCR amplification, for
PT diagnosis and drug development e.g. for allergic diseases and asthma -
XX
PS Claim 1; Page 18-31; 47pp; Japanese.
XX

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: January 6, 2002, 09:45:57 ; Search time 2642.04 Seconds
(without alignments)
97.613 Million cell updates/sec

Title: US-09-897-042-13
Perfect score: 24
Sequence: 1 atgaacttggtaccgatgaactaa 24

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_htc:*
10: gb_est1:*
11: gb_est2:*
12: gb_htc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		18.4	76.7	577	13	AQ453713	AQ453713 HS_5155_A
2	c	18.2	75.8	181	10	AA585156	AA585156 ATH637 HT
3		18.2	75.8	228	10	BB258711	BB258711 BB258711
4	c	18.2	75.8	495	10	AU221054	AU221054 AU221054
5	c	18.2	75.8	648	10	AI054495	AI054495 coau0001C
6	c	18.2	75.8	763	13	CNS0255R	AL181656 Tetraodon
7	c	17.8	74.2	502	10	AW017100	AW017100 EST272522
8	c	17.8	74.2	1039	13	CNS06DN3	AL394021 T7 end of
9	c	17.6	73.3	328	10	BB120380	BB120380 BB120380
10	c	17.6	73.3	370	10	AV622113	AV622113 AV622113
11	c	17.6	73.3	427	13	AZ144229	AZ144229 SP_0020_B
12	c	17.6	73.3	500	13	AZ173900	AZ173900 SP_0125_B
13	c	17.6	73.3	533	11	BF349591	BF349591 MR0-HT016
14	c	17.6	73.3	560	13	AZ794406	AZ794406 2M0048K10
15		17.6	73.3	562	13	AZ154963	AZ154963 SP_0040_A
16		17.6	73.3	593	13	BH022435	BH022435 GH_MBb000
17	c	17.6	73.3	677	11	BG856621	BG856621 1024047C0
18	c	17.6	73.3	682	11	BG848783	BG848783 1024022G0
19	c	17.6	73.3	764	11	BG848784	BG848784 1024022G0
20	c	17.6	73.3	809	11	BG784924	BG784924 SEAUMC004
21	c	17.6	73.3	921	13	AZ201725	AZ201725 SP_0055_A
22		17.6	73.3	946	13	CNS01OXD	AL154097 Anopheles
23	c	17.6	73.3	1045	13	CNS06OYH	AL408687 T3 end of
24	c	17.4	72.5	338	10	AA932931	AA932931 oo04h01.s
25	c	17.4	72.5	393	10	AA905887	AA905887 oj82e12.s
26	c	17.4	72.5	400	13	AQ505888	AQ505888 RPCI-11-2
27	c	17.4	72.5	428	11	T89283	T89283 yd37e07.s1
28	c	17.4	72.5	483	11	BG486615	BG486615 dd03a09.x
29		17.4	72.5	506	10	AW634839	AW634839 bl25f07.w
30	c	17.4	72.5	523	10	AW444944	AW444944 UI-H-BI3-
31		17.4	72.5	677	10	AV654637	AV654637 AV654637
32		17.2	71.7	303	10	BB202920	BB202920 BB202920
33	c	17.2	71.7	407	13	AQ033034	AQ033034 HS_2226_A
34		17.2	71.7	424	11	BG042989	BG042989 saa46g03.
35	c	17.2	71.7	467	10	BE111384	BE111384 UI-R-BJ1-
36		17.2	71.7	471	10	BE359101	BE359101 DG1_38_F0
37		17.2	71.7	479	11	BI075714	BI075714 IP1_24_C0
38		17.2	71.7	480	10	AU203508	AU203508 AU203508
39	c	17.2	71.7	518	10	AA409213	AA409213 EST03694
40		17.2	71.7	557	10	AJ280888	AJ280888 4A3A-AAY-
41		17.2	71.7	573	11	BG466085	BG466085 RHIZ2_48_
42	c	17.2	71.7	629	10	AW828984	AW828984 ra70e02.y
43		17.2	71.7	641	10	AW355303	AW355303 pnf-b.pk0
44	c	17.2	71.7	663	11	BG063351	BG063351 H3006B08-
45		17.2	71.7	709	13	AZ561191	AZ561191 RPCI-23-2

ALIGNMENTS

RESULT 1
AQ453713
LOCUS AQ453713 577 bp DNA GSS 21-APR-1999
DEFINITION HS_5155_A2_E05_SP6E RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=731 Col=10 Row=I, DNA sequence.
ACCESSION AQ453713
VERSION AQ453713.1 GI:4594883
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 577)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Resear h Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu
Plate: 731 row: I column: 10
Seq primer: SP6

```
          Class: BAC ends
          High quality sequence stop: 577.
FEATURES             Location/Qualifiers
     source          1. .577
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=731 Col=10 Row=I"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /sex="male"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
                     EcoRI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"
BASE COUNT      152 a    101 c    141 g    180 t      3 others
ORIGIN


  Query Match             76.7%;  Score 18.4;  DB 13;  Length 577;
  Best Local Similarity   95.0%;  Pred. No. 2.9e+02;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 tgaacttggtaccgatgaac 21
          |||||||||||||||||| ||
Db     53 TGAACTTGGTACCGATGGAC 72


RESULT   2
AA585156/c
LOCUS       AA585156      181 bp    mRNA            EST       09-SEP-1997
DEFINITION  ATH637 HTCDL1 Homo sapiens cDNA 5'/3', mRNA sequence.
ACCESSION   AA585156
VERSION     AA585156.1  GI:2385044
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 181)
  AUTHORS   Sohn,U., Park,D.S., Lee,C.M., Cho,W.K., Ahn,H.J., Lee,M.Y., Hwang
            ,M.Y. and Jin,S.W.
  TITLE     Human HTCDL1 library cDNAs
  JOURNAL   Unpublished (1994)
COMMENT     Contact: Uik Sohn,D.S.Park,C.M.Lee,W.K.Cho,H.J.Ahn,M.Y.Lee
            ,M.Y.Hwang,S.W.Jin
            Laboratory of Molecular Biology
            Kyungpook National University
            Dept. of Genetic Eng.,Kyungpook National Univ., Taegu 702-701,
            Korea
            Tel: 82-053-950-5382
            Fax: 82-053-955-5327
            Email: usohn@bh.kyungpook.ac.kr
            Seq primer: M13 Reverse/SK primer.
FEATURES             Location/Qualifiers
     source          1. .181
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="HTCDL1"
                     /lab_host="XL1-Blue"
                     /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
                     EcoRI; Poly(A)-mRNA from the 2-year old male fetal
                     thymus, oligo(dT) priming, EcoRI cloning in the vector
                     pBluescipt (Stratagene)."
BASE COUNT       65 a     21 c     22 g     73 t
ORIGIN


  Query Match             75.8%;  Score 18.2;  DB 10;  Length 181;
  Best Local Similarity   87.0%;  Pred. No. 3.2e+02;
  Matches   20;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 atgaacttggtaccgatgaacta 23
          |||||||||||| | ||||| ||
Db    168 ATGAACTTGGTAACAATGAAATA 146


RESULT   3
BB258711
LOCUS       BB258711      228 bp    mRNA            EST       06-JUL-2000
DEFINITION  BB258711 RIKEN full-length enriched, 7 days neonate cerebellum Mus
            musculus cDNA clone A730086A03 3', mRNA sequence.
ACCESSION   BB258711
VERSION     BB258711.1  GI:8951544
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 228)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
            ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
            Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
            Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
            Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.
            , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.
            , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.
            , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya
            ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.
            , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
            ,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al.)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
            ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
             Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
             Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
             Carninci,P. and Hayashizaki,Y.
             High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
             Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .228
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="A730086A03"
                     /clone_lib="RIKEN full-length enriched, 7 days neonate
                     cerebellum"
                     /tissue_type="cerebellum"
                     /dev_stage="7 days neonate"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: January 6, 2002, 09:47:47 ; Search time 1586.23 Seconds
(without alignments)
249.606 Million cell updates/sec

Title: US-09-897-042-13
Perfect score: 24
Sequence: 1 atgaacttggtaccgatgaactaa 24

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vi:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rod:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rod:*
36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	Query Match	Length	DB	ID	Description
c	1	19.8	82.5	57329	9	AL391649	AL391649 Human DNA
c	2	19.8	82.5	162149	2	AC023574	AC023574 Homo sapi
c	3	19.2	80.0	170343	2	AC015939	AC015939 Mus muscu
	4	19.2	80.0	174162	2	AC024611	AC024611 Mus muscu
	5	19.2	80.0	241880	10	AC027285	AC027285 Mus muscu
	6	19.2	80.0	243267	2	AC011407	AC011407 Homo sapi
	7	18.2	75.8	32962	3	CELR05H11	U00056 Caenorhabdi
	8	18.2	75.8	58949	3	CELK07E12	U00054 Caenorhabdi
	9	18.2	75.8	166536	2	AC018465	AC018465 Homo sapi
	10	17.8	74.2	29482	3	CEW05B5	Z82071 Caenorhabdi
	11	17.8	74.2	89004	3	CEY69H2	Z98877 Caenorhabdi
c	12	17.8	74.2	102433	8	AY013246	AY013246 Hordeum v
c	13	17.8	74.2	145576	8	AP002541	AP002541 Oryza sat
c	14	17.8	74.2	147857	8	AP002747	AP002747 Oryza sat
c	15	17.6	73.3	2540	8	SPPAP1	X57078 S.pombe pap
	16	17.6	73.3	4993	1	CAU17110	U17110 Clostridium
	17	17.6	73.3	11949	1	U32808	U32808 Haemophilus
c	18	17.6	73.3	13403	1	AE007768	AE007768 Clostridi
c	19	17.6	73.3	17487	3	DMC171E4	AL021726 Drosophil
	20	17.6	73.3	19469	8	SPAC1783	AL136499 S.pombe c
c	21	17.6	73.3	21693	2	AC017390	AC017390 Drosophil
	22	17.6	73.3	21838	1	PDCRAFOPER	L32093 Pediococcus
	23	17.6	73.3	21839	1	PPSURFOP	Z32771 P.pentosace
c	24	17.6	73.3	31144	2	AC087398	AC087398 Homo sapi
	25	17.6	73.3	51858	9	AF097649	AF097649 Homo sapi
	26	17.6	73.3	110000	2	AC092449_0	AC092449 Homo sapi
c	27	17.6	73.3	112363	2	AC087425	AC087425 Oryza sat
c	28	17.6	73.3	115861	9	AC006971	AC006971 Homo sapi
	29	17.6	73.3	123209	9	AL158195	AL158195 Human DNA
c	30	17.6	73.3	143107	9	AC005949	AC005949 Homo sapi
	31	17.6	73.3	145012	8	AC087551	AC087551 Oryza sat
c	32	17.6	73.3	149684	9	AL355485	AL355485 Human DNA
c	33	17.6	73.3	154065	9	AC005948	AC005948 Homo sapi
	34	17.6	73.3	156292	9	CNS05TD6	AL356016 Human chr
	35	17.6	73.3	193430	2	AC011709	AC011709 Homo sapi
	36	17.6	73.3	198127	2	AL512510	AL512510 Homo sapi
	37	17.6	73.3	210613	2	AC044907	AC044907 Homo sapi
c	38	17.6	73.3	296497	3	AE003421	AE003421 Drosophil
	39	17.4	72.5	885	9	AF143325	AF143325 Homo sapi
c	40	17.4	72.5	40118	9	HSU63313	U63313 Human cosmi
c	41	17.4	72.5	41935	9	HSU61B11	Z73913 Human DNA s
c	42	17.4	72.5	64356	9	HS512B11	AL031058 Human DNA
c	43	17.4	72.5	71990	2	AL590044	AL590044 Homo sapi
c	44	17.4	72.5	148434	2	AL133543	AL133543 Homo sapi
c	45	17.4	72.5	169949	9	AC084358	AC084358 Homo sapi

ALIGNMENTS

RESULT 1
AL391649/c
LOCUS AL391649 57329 bp DNA PRI 12-JAN-2001
DEFINITION Human DNA sequence from clone RP11-202E10 on chromosome 9 Contains GSSs and STSs, complete sequence.
ACCESSION AL391649
VERSION AL391649.8 GI:11878445
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 57329)
AUTHORS Sehra,H.
TITLE Direct Submission
JOURNAL Submitted (04-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT On Dec 18, 2000 this sequence version replaced gi:11876078.
During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated repeat sequence elements. Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:
Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr9
RP11-202E10 is from the library RPCI-11.1 constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
IMPORTANT: This sequence is not the entire insert of clone RP11-202E10 It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.
The true left end of clone RP11-477G9 is at 57230 in this sequence.
The true right end of clone RP11-33G16 is at 100 in this sequence.

FEATURES Location/Qualifiers

source 1. .57329
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-202E10"
/clone_lib="RPCI-11.1"

misc_feature complement(1. .100)
/note="match: STS: Em:G50491
match: GSS: Em:AQ046842"

misc_feature 110. .609
/note="match: GSS: Em:AQ834379"

repeat_region 1506. .1774
/note="L2 repeat: matches 2263. .2542 of consensus"

repeat_region 2279. .2402
/note="FLAM_A repeat: matches 2. .127 of consensus"

repeat_region 2409. .2705
/note="AluY repeat: matches 1. .297 of consensus"

repeat_region 2723. .2864
/note="BC200 repeat: matches 1. .142 of consensus"

misc_feature 3300. .3830
/note="match: GSS: Em:AQ666909"

misc_feature complement(4552. .4767)
/note="match: STS: Em:G04418"

repeat_region 4886. .5133
/note="AluSg repeat: matches 3. .290 of consensus"

repeat_region 5633. .5708
/note="38 copies 2 mer tt 65% conserved"

repeat_region 5714. .5910
/note="L1PB3 repeat: matches 5951. .6150 of consensus"

repeat_region 5911. .6273
/note="MSTA repeat: matches 1. .426 of consensus"

repeat_region 6274. .7925
/note="MSTA-internal repeat: matches 1. .1651 of consensus"

repeat_region 7926. .8304
/note="MSTA repeat: matches 1. .426 of consensus"

repeat_region 8305. .8814
/note="L1PB3 repeat: matches 5455. .5951 of consensus"

repeat_region 9190. .10603
/note="L1MA2 repeat: matches 4903. .6308 of consensus"

repeat_region 10604. .11035
/note="L1PB2 repeat: matches 5712. .6155 of consensus"

[redacted]peat_region 11036. .11809
/note="L1MA2 repeat: matches 4128. .4903 of consensus"

[redacted]region 14222. .14608
/note="HAL1 repeat: matches 397. .791 of consensus"

repeat_region 14791. .14891
/note="L1MC4 repeat: matches 7860. .7956 of consensus"

repeat_region 14900. .15211
/note="AluSc repeat: matches 1. .308 of consensus"

misc_feature complement(15008. .15360)
/note="match: GSS: Em:B37456"

repeat_region 16621. .16927
/note="AluSx repeat: matches 1. .312 of consensus"

repeat_region 16933. .17651
/note="L1MD repeat: matches 973. .1698 of consensus"

repeat_region 17783. .18059
/note="AluJb repeat: matches 1. .286 of consensus"

repeat_region 18689. .18937
/note="L1M4 repeat: matches 4078. .4330 of consensus"

repeat_region 19263. .19375
/note="L2 repeat: matches 2626. .2743 of consensus"

repeat_region 19497. .19983
/note="L1MA9 repeat: matches 5782. .6275 of consensus"

repeat_region 19984. .20035
/note="26 copies 2 mer ta 82% conserved"

repeat_region 20038. .20565
/note="L1MA9 repeat: matches 5272. .5789 of consensus"

repeat_region 22718. .22833
/note="L2 repeat: matches 2624. .2750 of consensus"

repeat_region 23588. .23895
/note="AluY repeat: matches 3. .308 of consensus"

repeat_region 24075. .24372
/note="AluJb repeat: matches 1. .303 of consensus"

repeat_region 24404. .24783
/note="190 copies 2 mer ta 61% conserved"

repeat_region 24418. .24781
/note="91 copies 4 mer tata 62% conserved"

repeat_region 24848. .25495
/note="L1PA13 repeat: matches 5494. .6144 of consensus"

repeat_region 27540. .27680
/note="MIR repeat: matches 21. .173 of consensus"

repeat_region 28201. .28287
/note="L2 repeat: matches 2585. .2673 of consensus"

repeat_region 28785. .28910
/note="L2 repeat: matches 2567. .2698 of consensus"

repeat_region 29043. .29139
/note="MER5A repeat: matches 10. .106 of consensus"

repeat_region 29356. .29540
/note="MER5A repeat: matches 7. .189 of consensus"

repeat_region 29760. .29981
/note="L2 repeat: matches 1736. .1957 of consensus"

repeat_region 29982. .30347
/note="THE1B repeat: matches 1. .364 of consensus"

repeat_region 30350. .31896
/note="THE1B-INTERNAL repeat: matches 1. .1578 of consensus"

repeat_region 31897. .32263
/note="THE1B repeat: matches 1. .364 of consensus"

repeat_region 32264. .32308
/note="L2 repeat: matches 1686. .1737 of consensus"

repeat_region 32554. .32867
/note="AluSq repeat: matches 1. .301 of consensus"

repeat_region 33940. .34108
/note="MER72 repeat: matches 1. .169 of consensus"

repeat_region 35073. .35486
/note="MSTA repeat: matches 1. .422 of consensus"

repeat_region 35810. .36001
/note="MIR repeat: matches 71. .262 of consensus"

repeat_region 36220. .36517
/note="AluY repeat: matches 1. .299 of consensus"

repeat_region 36537. .37220
/note="L1PA16 repeat: matches 5440. .6157 of consensus"

repeat_region 37216. .38182
/note="L1PA16 repeat: matches 4411. .5420 of consensus"

repeat_region 39220. .39333
/note="MIR repeat: matches 59. .172 of consensus"

repeat_region 39657. .39931